*American Economic Review 2022, 112(10): 3141–3175*
*https://doi.org/10.1257/aer.20190668*

# Persistent Overconfidence and Biased Memory: Evidence from Managers†

*By* David Huffman, Collin Raymond, and Julia Shvets*

*A long-standing puzzle is how overconfidence can persist in settings characterized by repeated feedback. This paper studies managers who participate repeatedly in a high-powered tournament incentive system, learning relative performance each time. Using reduced form and structural methods we find that (i) managers make overconfident predictions about future performance; (ii) managers have overly positive memories of past performance; (iii) the two phenomena are linked at an individual level. Our results are consistent with models of motivated beliefs in which individuals are motivated to distort memories of feedback and preserve unrealistic expectations.* (*JEL* D82, D83, J33, L25, L81, M52, M54)

Overconfidence has often been described as a fundamental bias in human decision making (e.g., Smith 1776). A long-standing puzzle, however, is whether and how overconfidence can be more than an ephemeral phenomenon. In many of the settings where economic theory posits a crucial role for beliefs about relative performance—the workplace, school, university, and competitive environments more generally—individuals receive repeated performance feedback, which would seemingly lead to the correction of overconfidence if there is Bayesian updating.

Economists have considered different mechanisms that might generate persistent overconfidence, but one leading explanation is "motivated beliefs." The idea is that individuals may be motivated to preserve unrealistic expectations, e.g., because they gain utility directly from optimistic beliefs, or because confidence helps provide

* Huffman: University of Pittsburgh (email: huffmand@pitt.edu); Raymond: Cornell University (email: collinbraymond@gmail.com); Shvets: Christ's College, University of Cambridge (email: js591@cam.ac.uk). Roland Bénabou was the coeditor for this article. We thank the editor and three anonymous referees for helpful comments, as well as Sandro Ambuehl, Ned Augenblick, Heski Bar-Isaac, Doug Bernheim, Matias Cattaneo, Gary Charness, Armin Falk, Yoram Halevy, Paul Heidhues, PJ Healy, Mitch Hoffman, Jun Ishii, Botond Kőszegi, Yiming Liu, George Loewenstein, Ulrike Malmendier, Luis Santos-Pinto, Jean Tirole, Lise Vesterlund, Leeat Yariv, Florian Zimmermann, and seminar and workshop participants at SITE; Santa Barbara Conference on Experimental and Behavioral Economics; Advances with Field Experiments; CESifo Behavioral Economics; Princeton; Carnegie Mellon; ESA, Tucson; Institute for Research on Behavior and Inequality, Bonn; University of Nottingham; University of Trento; University of Lund; Yonsei University; Hong Kong University; Middlesex University; Pennsylvania State University; University of Cambridge; and Vienna University of Economics and Business for helpful comments. We thank the Keynes Fund for financial support; Martin Oldfield for discussions, and help with logistics and payments; Sidney Ho for help processing payments. For excellent research assistance, we thank Mihir Bhatia, Marta Boczon, Lukas Hensel, Balint Lenkei, Myrto Oikonomou, and Lu Wang. We are grateful to the firm for providing us with the data and to Nick, Eliana, Stephen, and Eric for all their help.

† Go to https://doi.org/10.1257/aer.20190668 to visit the article page for additional materials and author disclosure statements.

motivation in the face of self-control problems; for a survey, see Bénabou and Tirole (2016). Models of motivated beliefs make various assumptions about how individuals are able to sustain overconfidence, for example assuming that individuals can use a technology of biased memory to selectively distort memories of past feedback; basing predictions on overly positive memories, future selves will be overconfident. A motivated beliefs explanation for overconfidence has important implications, because overconfidence may persist in the face of feedback, individuals may not respond to information in standard ways, and the welfare effects of overconfidence are ambiguous and can even be positive.

This paper seeks to establish (i) whether there is persistent overconfidence about relative task performance in an important field setting—a workplace in which managers compete regularly for performance bonuses while receiving detailed feedback, (ii) whether these managers have overly positive memories about past workplace performance, and (iii) whether overly positive memories are associated with making overconfident predictions. Our results are affirmative on all three questions, and provide, to our knowledge, the first evidence of persistent overconfidence about the future being linked to biased memories of the past. This is in line with explanations for overconfidence based on motivated beliefs.

Section I of the paper describes the firm, our study design, and the data. The study involved approximately 230 managers, each of whom runs a separate store. The managers compete repeatedly in a high-powered tournament incentive scheme, with detailed feedback, and many managers have observed a large number of tournament outcomes. One source of data is the historical records of the firm on each manager's tournament outcomes. The other is a lab-in-the-field study. This study elicited manager predictions about relative performance in the upcoming tournament at their job, for 2015:IV (quarter four of 2015), as well as memories about performance in a previous tournament, in 2015:II, for which results had been provided approximately two months earlier.

Our data allow us to address some of the challenges that studies face when trying to establish overconfidence bias. One key issue for assessing whether predictions are reasonable is the need to take into account what information individuals had access to when making predictions. For example, predictions of the future may "appear" overconfident ex post, relative to realized outcomes, but in fact be fully Bayesian if one takes into account the signals that informed these beliefs ex ante (see Benoît and Dubra 2011; Benoît, Dubra, and Moore 2015). Oftentimes, researchers do not have any information about past signals, but in our setting, a key type of signal, past tournament outcomes, is public and observable. We can thus check directly whether manager predictions are explainable by ex ante public signals. We also assess in various ways whether the results could be explained by managers having access to additional, private signals.

Section II of the paper presents a reduced form analysis that speaks to our three main research questions. First, we show that managers are overconfident relative to a range of different reduced form predictors one could form using past public signals. For example, 48 percent of managers are overconfident relative to the predictions of a panel regression model that takes lagged tournament performances as predictors, compared to only 21 percent being underconfident, and the average prediction is overconfident by about 0.5 quintiles. This overconfidence is similarly prevalent and large among managers with substantial experience, so overconfidence is persistent in the face of feedback. This latter result also casts doubt on an explanation based on some forms of private signals, e.g., ones received before the job, that lead to overconfident priors. If managers are Bayesian, such overconfidence should disappear as they observe more tournament outcomes. Second, our analysis provides evidence consistent with managers being motivated to have positive memories of past performance. Specifically, top-performing managers are quite accurate in recalling their good performances. Managers with performances below the very top, however, have substantial recall errors, and these are strongly skewed toward overly positive memories. Third, our analysis shows that it is those managers who have overly positive memories of past signals who are particularly likely to make overconfident predictions about future performance, so these phenomena are linked in a way predicted by models of motivated beliefs.

Section III offers a complementary structural analysis that allows us to go beyond the reduced form analysis in several important ways. First, we can formulate a model of Bayesian learning in our setting, and assess whether managers are overconfident relative to this explicit Bayesian benchmark. The findings mirror those from our reduced form analysis, in that 45 percent of managers are overconfident relative to what the structural model says they should have predicted, versus only 26 percent underconfident, and we can reject statistically that the model matches the data. Second, while one can think of forms of private signals that are not ruled out by our reduced form analysis, we can discipline such explanations with an extension of the structural model, which explicitly models private signals, and allows these to take whatever form best fits the data. It is not necessarily the case that the best fitting model will come close to the data, however, because the information about priors based on public signals puts restrictions on how much overconfidence can be rationalized by private signals. The results show that the best-fitting signal structure has features that seem quite implausible, and furthermore, the model is far from matching the data based on standard confidence intervals. Third, we can extend the structural model to incorporate biased memory of past tournament outcomes, in a way that is disciplined by our data. Whereas our reduced form analysis tests a qualitative prediction, that overconfidence should be positively correlated with biased memory, we can use the structural model to ask whether biased memory can explain the data in quantitative terms: predicting who is overconfident, and to what extent. The model with biased memory comes closer to the data than other versions of the structural model, and we cannot reject that this extension matches the data statistically. Taken together, our findings are consistent with explanations for persistent overconfidence provided by models of motivated beliefs.

In Section IV we discuss the implications of our findings, and also provide some additional exploratory analysis on how manager overconfidence relates to performance and management style. The latter analysis shows that overconfident managers do not perform any worse than other managers, but exhibit some differences in management style. Specifically, overconfident managers tend to hire fewer assistant managers in their stores, and give less discretion to employees in a lab-in-the field experiment on management style and delegation.

The results of this paper speak to the empirical relevance of a theoretical literature on motivated beliefs. Models in this literature assume that under certain conditions individuals can have a motivation to "demand" confidence.[1] This can reflect a direct utility benefit from positive beliefs, due to self-esteem, "ego-utility," or anticipatory utility reasons (Bénabou and Tirole 2002; Kőszegi 2006; Brunnermeier and Parker 2005; Bracha and Brown 2012; Sarver 2018), or because confidence can help individuals work harder in the future (Bénabou and Tirole 2002; Compte and Postlewaite 2004). In models of motivated beliefs individuals also have some way of "supplying" distorted beliefs, albeit typically subject to some "reality constraints" that limit how far individuals want to, or are able to, distort beliefs away from the truth. For example, one strand of the literature assumes that individuals can use a technology of memory distortion (at a cost) to distort memories of past signals in an overly positive direction. This can be used to foster overconfidence because future selves will base beliefs on falsely positive memories (Bénabou and Tirole 2002; Compte and Postlewaite 2004; Gottlieb 2010).[2] Individuals might also be motivated to distort memory because of a direct utility benefit of positive memory, in the spirit of ego utility models like Kőszegi (2006). Importantly, because of how overconfidence and overly positive memories are linked in models of motivated beliefs (causally or due to a common underlying motivation), the models predict that overconfidence will tend to go hand in hand with overly positive memory. The evidence in this paper is consistent with this signature prediction of a motivated beliefs explanation for overconfidence, although it does not rule out that other factors might also play a role.[3]

The paper also contributes to an empirical literature on overconfidence in field settings. For example, the behaviors of investors, CEOs, gym members, and others (Barber and Odean 2001; Malmendier and Tate 2005, 2015; DellaVigna and Malmendier 2006; Oster, Shoulson, and Dorsey 2013; Cheng, Raina, and Xiong 2014), and the beliefs of truckers and professional poker players (Hoffman and Burks 2020; Park and Santos-Pinto 2010), all show signs of overconfidence in their respective decision environments, even though these individuals are presumably observing signals that should challenge their beliefs. Besides studying managers, and taking different approaches to address confounds such as private information, our paper is distinct because it shows a link between overconfidence and biased

[1] Relevant conditions include details of the decision environment, such as whether outcomes are relevant for self-esteem, or whether effort and ability are substitutes or complements. We discuss the plausibility of these conditions for our setting in Section I.

[2] Other technologies considered in the literature include taking steps to limit exposure to negative feedback (e.g., Carrillo and Mariotti 2000; Kőszegi 2006), self-signaling, i.e., taking actions or stating beliefs that signal confidence to future selves (Quattrone and Tversky 1984; Bénabou and Tirole 2004; Mijovic-Prelec and Prelec 2010; Bénabou and Tirole 2011). Some models assume individuals directly choose beliefs about themselves or future outcomes (e.g., Brunnermeier and Parker 2005).

[3] For example, there could also be a role for cognitive mistakes in generating manager overconfidence. Hoffman and Burks (2020) discuss the possibility that individuals have biased priors, but also underestimate the informativeness of all types of signals, thereby slowing learning. The Dunning-Kruger effect, discussed in social psychology, is similar in spirit in that low ability people fail to understand their incompetence (Kruger and Dunning 1999). A different type of explanation is that individuals have priors that put zero probability on having low ability; no amount of signals can cause a Bayesian to update a prior of zero to a positive probability (Heidhues, Kőszegi, and Strack 2018; see also Hestermann and Le Yaouanq 2021). Some other explanations for (potentially temporary) overconfidence assume overconfident priors, which might or might not be motivated, e.g., Santos-Pinto and Sobel (2005); Van den Steen (2004).

memory, which in turn points to an explanation for persistent overconfidence based on motivated beliefs.

Finally, the paper complements an empirical literature on overconfidence and motivated beliefs in the laboratory. Many studies have measured apparently overconfident behavior (e.g., Camerer and Lovallo 1999), and some have demonstrated overconfident beliefs using designs that can rule out any Bayesian explanation (Merkle and Weber 2011; Burks et al. 2013; Benoît, Dubra, and Moore 2015). Lab evidence on motivated beliefs includes Eil and Rao (2011) who find that individuals adjust beliefs more in response to good than bad information, immediately after it is received, and Schwardmann and van der Weele (2019) who find that subjects become more overconfident when they have a strategic need to impress others.[4] Related to biased memory, Chew, Huang, and Zhao (2020) show evidence that students can have falsely positive memories of performance on a cognitive ability test, and Zimmermann (2020) shows that memories of feedback about a cognitive ability test are accurate immediately after feedback but are biased one month afterwards. The lab evidence has clear strengths in terms of control and causality. Our study is complementary by providing evidence that the mechanisms of motivated beliefs and biased memory are empirically relevant when it comes to remembering and predicting real workplace performance. Also, previous studies documenting biased memory have not looked at overconfidence bias; ours is the first paper to test directly for a link between these phenomena.

## I. Work Setting and Datasets

### A. *Nature of the Work Setting*

The subjects of the study are managers working for a chain of food and beverage stores in a developed country. Each manager is in charge of a separate store, and makes a range of important decisions: the number of workers to employ, task allocation, and how many and which types of products to sell. A typical store has roughly 14 employees including one or more assistant managers. The manager receives a base salary, but can also earn substantial performance bonuses, based on his or her rank in a tournament conducted each quarter.

### B. *Incentive Scheme*

The tournament incentive scheme is intended by the firm to reward managers for hard work and ability, with better managers receiving better ranks.[5] A manager's rank in the quarterly tournament is determined by relative performance on

[4] There are mixed results on whether people engage in asymmetric updating about good versus bad news. See, e.g., Mobius et al. (2011); Barron (2021); Coutts (2019); and Schwardmann and van der Weele (2019). For other types of evidence on motivated overconfidence, see Charness, Rustichini, and Van de Ven (2013) and Hoffman (2016). There is also evidence for motivated beliefs in the domain of prosociality, with individuals desiring to believe that they are a prosocial person (see Haisley and Weber 2010; Gneezy et al. 2015; Di Tella et al. 2015; Carlson et al. 2020).

[5] The firm cares not just about incentivizing high effort, but also about rewarding high ability, because it wants to retain talented managers. The firm has used a version of the scheme for several decades, a sign that it views the scheme as successful at rewarding good managers.

four dimensions: (i) a measure of store profits that is designed to isolate manager contributions independent of store characteristics and location;[6] (ii) sales growth; (iii) a customer service rating by an undercover "mystery shopper"; (iv) an evaluation of the store manager by a regional manager against centrally set criteria.[7] A manager's position in the distribution for a given dimension puts him or her into one of several bands, with each band being assigned a score. The score values increase approximately linearly going from the worst to best band. The firm then multiplies the scores from the different dimensions to yield an aggregate performance measure denoted the base bonus. This is then multiplied by some extra factors—a group-based metric denoted area bonus, an extra factor for top-performing managers denoted top performer bonus, and an extra bonus factor for all stores—to yield a manager's final bonus score.[8] Finally, manager rank in the tournament is determined by ranking the final bonus score (with tiebreaking), with the best score being assigned rank 1.

The monetary amount of the performance bonus is calculated by multiplying the final bonus score by 30 percent of the base salary. The bonus thus rises continuously with rank and all ranks receive a prize. Because of the top performer bonuses there is convexity at the top of the scheme. Figure A1 in the online Appendix shows the shape of the incentive scheme. Managers get a substantial portion of earnings from the scheme. The median bonus is equal to about 22 percent of the base quarterly salary. The strength of incentives, in terms of the prize spread, is also substantial. The median bonus for the best (fifth) quintile of performance is about 36 percent of the base quarterly salary, compared to only about 13 percent for the bottom quintile. A more "local" measure of the strength of incentives is the reduction in earnings from dropping by one quintile. Reflecting convexity, this is 8 percent of the base quarterly salary going from quintile 5 to 4, and about 4 percent for each of the quintiles 1 to 4.

### C. *Communication and Feedback to Managers*

The firm's communications to managers emphasize that managers can influence their outcomes in the tournament, and try to foster pride and self-esteem in good tournament rankings. In one of the main internal communications to managers, for example, the firm notes that by understanding the ranking scheme and concentrating on this, a manager can "influence how much you earn each quarter." The firm links tournament rank to self-esteem by describing it as the key overall metric of being a good manager, by describing the bonus as a reward, by emphasizing that the job is challenging and requires skill, by how it describes top performers, and also through other means such as holding special parties to honor highly ranked managers.

[6] Profits are measured relative to targets constructed as predicted values from regressions of historical store profits on store characteristics such as region, store age, etc.. Managers in stores with more favorable characteristics or locations thus face higher profit targets.

[7] The review by a senior manager evaluates adherence to, for example, health and safety rules.

[8] Specifically, the Area Bonus is determined by averaging performance of the manager's store on each of the four dimensions with the performances of other stores in the nearby geographic area (there are typically between 5 to 10 stores in an area), then ranking performance of the area relative to performances of other areas on that dimension, and then splitting the ranking into bands each associated with a score value. These scores are multiplied across the four dimensions to yield the Area Bonus score. Top Performer bonuses are assigned to roughly the top 20 managers in the distribution. The extra bonus factor for all stores is based on performance of the company as a whole.

The firm also gives managers detailed feedback about their performance every quarter. Feedback comes in the form of a table, received by each manager, known as a "ranking table." In line with the central importance and salience of the overall rank as a performance metric, the first column in the table is about rank, giving the complete ranking of managers in the tournament. Subsequent columns give information about the various submetrics that determine rank (absolute and relative scores on the individual dimensions, base bonus, area bonus, final bonus, etc.). Figure C1 in the online Appendix shows an example of how the table looks. Managers discuss the quarterly feedback with senior managers in regularly scheduled meetings after each quarter. Thus, managers do receive this feedback each quarter, although they can potentially forget the information later on.

### D. *Historical Performance Data*

The company has shared its historical data on manager performance from 2016:I going back to 2008:I. The data include overall performance, performance on each of the dimensions that underlie the aggregate performance measure, and a few pieces of additional information such as how many assistant managers the manager chooses to hire. Online Appendix B discusses additional details about the creation of the dataset. For example, we discuss how the scope of the tournament has varied across some quarters—nationwide in some quarters, but divided into a few large, regional tournaments in others—and how we construct an exactly comparable measure of performance over time.[9] The analysis checks robustness to including or excluding quarters with regional tournaments.

The historical performance data yield some descriptive statistics about the work environment and the managers. The average number of stores active in any given quarter over the sample period is about 230, but the company has grown over time, reaching about 300 stores by the end of the sample period. Managers sometimes switch stores during their tenure. For example, among managers working in 2015:IV, which is the quarter for which we elicited manager predictions, roughly 48 percent have switched stores at least once during their tenure. Median tenure in the current store is five quarters, and median total tenure at the company is ten quarters. Over the sample period the fraction of managers leaving the managerial job is around 6 percent per quarter, with no significant time trends.

The data also shed some light on the determinants of performance in the tournament, showing that managers matter for tournament outcomes, although store characteristics matter as well. For example, looking at managers who switch stores, a 1 standard deviation increase in the mean of a manager's performance at his or her past stores is associated with an increase of about 0.30 standard deviation in performance at a manager's current store, controlling for store characteristics.[10] Another

[9] Rank in a regional tournament is a good proxy for nationwide rank, i.e., rank if there had been a national tournament instead.

[10] This result is based on regressing current store performance on mean performance at past stores for managers who are present in 2015:IV and who have switched stores at least once. By controlling for store characteristics—store age, proxy for store size, train station location, indicators for 38 geographic areas—we help rule out that managers have similar performance over time because the current store has similar characteristics to previous stores (results available upon request). We also find that observable store characteristics for the current store in 2015:IV, and the most recent previous store, are largely uncorrelated, suggesting that the firm's assignment policies for

indication that managers matter is the tendency for new managers to have worse performance initially and improve over the first couple of months. This is consistent with a role for manager ability or skill in contributing to performance, with managers improving this trait over time on the job. This tendency has caused the firm to adopt a policy of excluding new managers from the regular tournament in their first quarter of tenure, and instead award bonuses based on easier metrics. Our empirical analysis excludes a manager's initial quarter at the firm.

### E. *Data on Manager Predictions, Memories, and Traits*

Measurements of manager confidence about future performance and memories about past performance were obtained in a lab-in-the-field study conducted with managers in early 2015:IV. To conduct the study, researchers attended a type of regularly occurring meeting organized by the firm, in which groups of roughly eight to ten store managers meet with a more senior manager. These meetings took place in private rooms in various locations, e.g., in store break rooms.

The study followed a standardized protocol across sessions (meetings in which the study took place). Managers were seated at a table with dividers between them, and were not allowed to speak to one another, to ensure that decisions were made individually. The study materials were provided in written form, but there was also a verbal summary of the instructions for each part by the attending researcher to ensure understanding (a researcher attended every session), and verbal instructions followed a script to ensure exactly the same delivery of information across sessions. Piloting with a few managers before the study made clear that the instructions needed to be very simple and clear, as the managers were not used to participating in such exercises.

To address potential manager concerns about confidentiality, the researchers conducting the sessions gave their academic affiliations, explained that they were not employed by the firm, and guaranteed that the managers' individual responses would be kept completely confidential from the firm and coworkers. It was also emphasized that funds came from an academic grant and that checks would be mailed directly to the managers' home addresses, by the researchers, early in 2016:I. Thus, no one in the company would ever learn the managers' individual earnings in the study.

A total of 239 managers participated in the study. About 56 percent were female, median age was 36, and median tenure at the company was 2.5 years. Managers received a participation payment of about \$20. The study was divided into ten parts that involved incentivized choices, with one randomly selected to be paid; on average managers earned roughly an additional \$20 in incentive payments from the study. There were 32 sessions, with the earliest taking place on October 22, 2015 and the latest taking place on December 7, 2015. Of the 32 sessions, 22 took place in October. This distribution of sessions over time lead to variation in how long ago managers had seen the tournament results they were asked to remember, and how

managers who switch do not involve assigning managers systematically to the same type of store over time. Results of fixed effects regressions are also consistent with managers mattering for performance; adding manager fixed effects to a regression of performance on store fixed effects doubles the adjusted $R^2$ and the manager fixed effects are jointly statistically significant ($p < 0.01$).

far in the future were the tournament results they were asked to predict. The analysis therefore investigates whether the timing of sessions is related to the accuracy of manager memories and quality of manager predictions.

*Measure of Managers' Predictions of Future Performance*.—The lab-in-the-field study elicited managers' predictions for how they would rank in the upcoming (nationwide) tournament for 2015:IV. We focused on eliciting manager predictions about rank because this is the central performance metric in a manager's work life, and because feedback about rank is particularly salient.

Managers were presented with a table with five rows, with each row corresponding to a quintile, and were asked to guess whether they would be in the top 20 percent, the second 20 percent, the third 20 percent, the fourth 20 percent, or the lowest 20 percent of the tournament ranking by ticking a box in the corresponding row (the top row was the best quintile). In other words, managers were asked for their modal quintile, based on their beliefs about the probabilities of different quintiles. The study provided an incentive to guess correctly: about \$22 for getting it right. The managers knew that researchers would check the outcomes of the tournament, once they were available, and then mail payments in 2016:I. See online Appendix D for instructions for the prediction measure.

Since our goal is to accurately measure manager beliefs, several features of the design were intended to minimize measurement error. The elicitation of predictions in a proctored (workplace) setting, without distractions, and the combination of both written and verbal instructions, was intended to minimize error due to inattentiveness or lack of understanding. The incentives provided in the study were also designed to enhance attention. Furthermore, managers arguably already had substantial incentives to overcome costs associated with thinking carefully about the tournament, due to the large amount of money tied up in the performance bonus scheme. Another source of measurement error would be if managers deliberately misrepresented their beliefs to impress others. The confidentiality protocol for our study, however, should have minimized motives to state false beliefs in order to impress coworkers or the employer. Managers could still try to impress the researchers with their responses, but providing incentives for correct guesses is the standard remedy in experimental economics for minimizing such motives. Furthermore, it is not clear that managers would expect researchers to be impressed if they state confident beliefs that are subsequently checked and verified to be wrong.[11]

[11] If managers habitually misrepresent their beliefs in their actual work setting, to impress coworkers or the firm, it is conceivable that they might habitually misrepresent beliefs in our study, despite the confidentiality protocol. To discuss whether this is plausible it is useful to distinguish between two types of signaling motives. The first would be managers trying to signal *confidence*, i.e., private information about high ability, by stating high beliefs. This is unlikely to be a viable strategy in the work setting, however, since there is very rich public information about everyone's past performance. A second motive would be trying to signal *overconfidence bias*, if this is viewed as a favorable trait; given the rich public information about past performance, managers could signal overconfidence by making overly optimistic predictions. For such signaling to be possible, however, there seemingly need to be at least some truly overconfident types among the managers, since signaling can only induce a belief in types that are of nonzero measure. Thus, this type of signaling would itself suggest that overconfidence bias is present among managers. There are some empirical reasons to doubt that overconfidence bias is viewed as a favorable trait in the workplace, however, since we do not find that overconfident managers perform better than other managers (see discussion at the end of the paper).

A different type of measurement error could arise because the study did not elicit complete probability distributions from managers, i.e., the likelihoods that they attached to ending up in each of the five quintiles. This was dictated by the need to keep the elicitation as simple and naturalistic as possible. Piloting suggested that more complex approaches, and the relatively complex rules needed to make responses incentive compatible, would not be well understood. The key benefit of this approach is we are confident that the managers understood what they were being asked. One potential downside of this elicitation approach is that risk averse individuals might want to insure themselves against poor performance on the job, by placing their bet on a low quintile in our prediction measure. Any such hedging (i.e., insurance) motives, however, would work against finding overconfidence. Thus, if managers engage in hedging, this source of measurement error makes our findings of overconfidence a lower bound. We also check whether predictions are related to a measure of manager risk aversion, and find no statistically significant relationship, which casts doubt on an insurance motive.

Models of motivated beliefs can predict that managers truthfully report beliefs in our measure that are systematically overconfident (relative to what is justified given their ex ante information), under certain conditions that are not implausible for our setting.[12] Relevant conditions for ego utility to play a role in generating overconfidence include (i) rank depends on managers and not just external factors; (ii) rank is seen as praiseworthy and relevant for self-esteem. As discussed in our description of the work setting, both of these conditions apply in our case.[13] If overconfidence is instead motivated by a desire to motivate future selves, a possibility modeled in Bénabou and Tirole (2002), a necessary condition is that effort and ability are complements. While complementarity is difficult to verify directly in our setting, it seems plausible. One reason is that the convex incentive scheme fosters complementarity: all else equal, higher beliefs about ability translate into more optimistic expectations about rank, which is associated with a higher marginal benefit of effort due to convexity of the scheme.[14] Models in which individuals are motivated by anticipatory utility can also predict overconfidence about rank, under the condition that disutility of "disappointment" when tournament outcomes are finally realized is outweighed by anticipatory utility of overly optimistic interim beliefs during the tournament.[15]

[12] If overconfidence is prevalent then we could observe managers making predictions that are overly optimistic on average. By contrast, standard (classical) forms of measurement error, due to factors such as inattention, would imply mean-zero prediction errors.

[13] To the extent that ego utility is based on beliefs about having high ability or skill, rather than beliefs about strong work ethic or high effort, the needed assumption for ego utility models to predict overconfidence can be further refined: it should be the case that managers influence tournament rank at least partly through ability and not just effort. As discussed in our description of the work setting, it seems that both skill and effort must matter to some extent.

[14] Our setting also resembles situations described in Bénabou and Tirole (2011) as featuring complementarity of effort and ability. If effort and ability are instead substitutes, such models predict underconfidence.

[15] Otherwise such models could predict "defensive pessimism" and thus underconfidence about rank. One possible explanation for the minority of managers we find who are underconfident is heterogeneity in disappointment aversion. Notably, a process of memory distortion could be one way that individuals might minimize although not entirely avoid a negative impact of disappointment; rather than remembering disappointments, individuals come to recall something more positive, which in turn fosters positive expectations for the future.

*Measure of Managers' Memories of Past Performance*.—Another part of the study asked managers to recall their rank in the most recent (nationwide) tournament, which was 2015:II. Managers had learned the results of this tournament roughly two months earlier. Specifically, managers were asked to recall their rank and offered a payment of \$1.50 for being within $+/-$ 10 ranks of their true past rank. The incentives provided for recall were smaller than in the prediction task, because recalling a number that they had learned and discussed with a superior is arguably easier than predicting the future. The instructions provided the header row of the tournament outcome table from the second quarter, and circled the relevant column header, to maximize clarity about what was being asked. Managers had to answer the question on the spot, and could not talk to each other or use their phones to look it up, so the question was a test of their memory. See online Appendix D for the instructions for the memory measure. The study also asked managers to remember second-quarter performances on some of the submetrics that determined their rank; we discuss these in robustness checks on the memory analysis.

Similar to the prediction measure, our design was intended to minimize measurement error due to inattention or due to managers misrepresenting their memories for some reason. The use of financial incentives, and a distraction-free environment for elicitation, were intended to foster attentiveness and reduce noise. The fact that substantial workplace incentives are tied to the performance indicator that we ask the managers to recall arguably implies that they should be willing to pay cognitive costs of recalling rank; we are asking them to recall something that is viewed as valuable information in the workplace. Financial incentives, and confidentiality, were intended to minimize any motives managers might have to overstate their recalled performance to impress coworkers or the researchers. Also, it is unclear that being inaccurate in recall is something that is viewed as impressive. Unlike for the prediction measure, there was no hedging motive for the memory measure, as it was retrospective rather than prospective.

Models of motivated beliefs provide a reason why manager responses to our memory measure might reflect truly inaccurate memories. These models imply that managers could value remembering a good rank in 2015:II, even if this deviates from the truth, under similar conditions that lead such models to predict overconfidence.[16] If this tendency is prevalent then the average recall error could involve remembering substantially better than actual performance, subject to some potential "reality constraints" that bound how much memory can be distorted. Specifically, the models predict that (i) managers with the best ranks in the second quarter should have accurate memory, because they cannot remember anything better, and there is no motive to remember worse; and (ii) managers who are below the top of the performance

[16] For example, managers could have ego utility from (potentially falsely) positive memories of past rank, under the condition that good rank is diagnostic of manager performance and relevant for self-esteem, which is plausible in our setting. A corollary is that managers may be less motivated to distort memory of performance metrics that are less tied to individual performance, a prediction that we explore in robustness checks for our analysis of manager memories. Alternatively, individuals who receive negative signals could be motivated to implement falsely positive memories for instrumental reasons: to shore up the confidence of future selves, and induce higher effort to overcome self-control problems (under the condition that effort and ability are complements). Motives related to anticipatory utility could also seemingly provide a reason to value positive memories of past rank, as long as disutility from negative surprises is not too strong, because these foster expectations of good future performance, and thus generate positive anticipatory utility.

distribution, by contrast, may have inaccurate memory, and these errors should be asymmetric in the direction of remembering better than actual performance.[17]

Models of motivated beliefs also make a prediction about the relationship between our memory and prediction measures at the individual level. The conditions that cause these models to predict overconfidence also cause them to predict overly positive memories. Thus, the models imply that if we observe a manager making overconfident predictions we should also tend to see them having overly positive memories, i.e., we should observe a positive correlation between overconfidence and overly positive memory.[18]

*Measures of Other Manager Traits*.—The study also measured some other manager traits in case these might be related to overconfidence: gender, as previous studies have found gender differences in overconfidence for some types of tasks (e.g., Niederle and Vesterlund 2007); experience at the company (tenure), to allow investigating whether greater exposure to feedback might be related to accuracy of predictions; and manager age, in case greater life experience is related to reduced overconfidence. These traits are featured in the main analysis as control variables. The study also included incentivized measures of willingness to misreport information, willingness to work on an addition task, knowledge and understanding of details of the firm's incentive scheme, and risk aversion. Nonincentivized measures include manager self-assessments of willingness to take risks, willingness to compete, relative confidence, and patience (more information about these control variables is provided in online Appendix E). We show in robustness checks that controlling for these does not change our results. The study also included an experiment designed to measure one potential aspect of an overconfident management style, unwillingness to delegate; this is used as an outcome variable in the exploratory analysis on management style discussed at the end of the paper. The study also had some additional measures of manager memories about the submetrics that determined overall rank in 2015:II, which we discuss briefly in our analysis on memory, and more extensively in online Appendix K. Finally, the study involved some measures that are not used in our analysis (online Appendix U gives the full set of instructions for the lab-in-the-field study).

## II. Reduced Form Analysis

### A. *Descriptives on Manager Predictions and Empirical Strategy for Identifying Overconfidence*

As a first look at the data, panel A of Figure 1 shows the distribution of manager predictions. The most salient feature is the skew toward predicting higher quintiles (throughout the paper we order quintiles such that 5 is the best). Only about

[17] With standard (classical) measurement error due to inattention, or imperfect but unbiased memory, one should expect the average recall error across managers to be zero: specifically, for the best performing managers to have downward errors, the worst performing managers to have upward errors of a similar size, and for errors in the middle of the distribution to be symmetric rather than skewed toward better than actual.

[18] Standard measurement error in the prediction and memory measures would not predict that the direction of errors should be positively correlated across the measures.

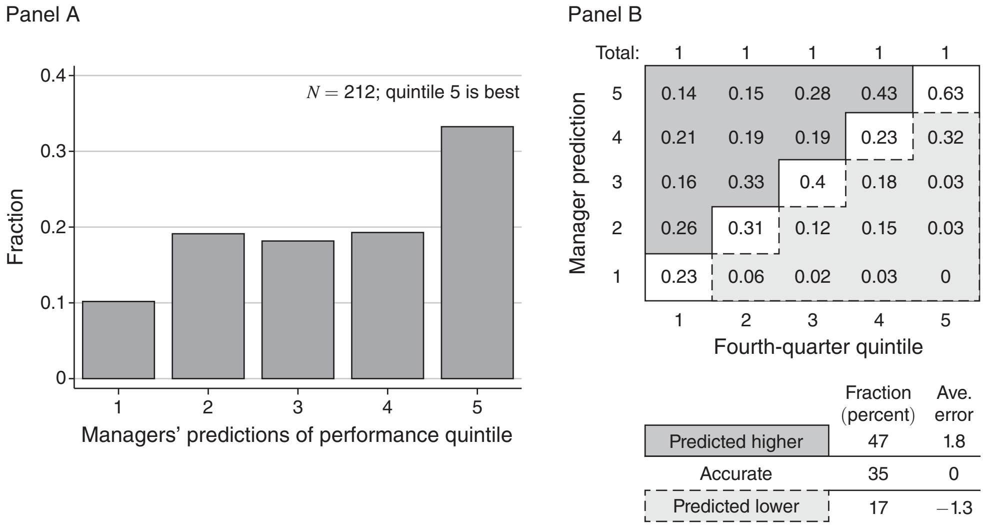


Figure 1. Distribution of Manager Predictions about Fourth Quarter and Comparison to Fourth-Quarter Realizations

10 percent predict achieving the worst quintile, roughly 20 percent predict each one of the intermediate quintiles, and 33 percent predict achieving the top quintile.

A comparison to realized outcomes in the fourth quarter, shown in panel B, suggests that managers do have insights into predicting future performance: achieved outcome and prediction are significantly positively correlated, 0.47 (Spearman; $p < 0.001$). On the other hand, panel B shows that managers make ex post prediction errors, and these errors are asymmetric: 47 percent of managers bet on a higher (better) quintile than their realized quintile, versus less than half as many, 17 percent, betting on a lower quintile. In terms of magnitudes, the errors are larger in the optimistic direction: 1.8 quintiles conditional on predicting higher than the realization, versus 1.3 quintiles conditional on predicting lower. On average managers predict a performance that is about 0.60 quintiles better than the realized quintile, a difference that is statistically significant from zero ($p < 0.001$).[19]

Although the skewness of manager predictions toward the best quintile in panel A goes in the direction of overconfidence, this is not sufficient, on its own, to establish overconfidence bias. As pointed out by Benoît and Dubra (2011), inferring overconfidence from bets on the mode can be very misleading, without information about the signals that individuals observe ex ante. Indeed, Benoît and Dubra (2011) show that in this case it is possible to rationalize almost everyone betting that their modal quintile is the best quintile. For example, if individuals have flat priors and the private signal structures characterized by frequent, weakly positive signals, then almost everyone can end up thinking they are slightly more likely to be in the best quintile than lower quintiles.[20]

[19] This result is from an ordinary least squares (OLS) regression of the prediction error on a constant term.

[20] Benoît and Dubra (2011) do not make claims that such information structures are generally plausible, but they point out the importance of taking into account the past signals that individuals have seen, for identifying overconfidence bias. This problem can be mitigated by eliciting the full probability distribution of beliefs about the likelihoods of all five quintiles, rather than just asking for the modal quintile as we did. With this information on intensity of beliefs, it is possible to test the Bayesian model by checking whether the average of posterior distributions across

The ex post prediction errors shown in panel B are also not sufficient to establish overconfidence bias, for similar reasons. Given flat priors, the appropriate private signals could lead many managers to predict high quintiles, and given randomness in tournament outcomes, ex post errors are to be expected even if managers are fully Bayesian. Burks et al. (2013) propose a statistical test that can assess whether ex post prediction errors are too extreme to be explained by the Bayesian model, even in the absence of information about private signals. By leveraging the additional information contained in realizations (and assuming that realizations reveal true types), the test imposes tighter restrictions than in the case of Benoît and Dubra (2011).[21] Nevertheless, ex post errors must still be relatively extreme to allow rejecting the model, because unobserved private signals give the model substantial flexibility. Applying their test to the ex post prediction errors of our managers, we cannot statistically reject the Bayesian model.[22]

While the evidence in Figure 1 does not by itself establish overconfidence bias, we can augment our observations of manager predictions with the historical performance data, and then test for overconfidence bias in a different way. The historical data include what are arguably the most important ex ante signals that managers should have used to form predictions: past tournament outcomes. This gives us information about the nonuniform priors managers should have had about their modal quintile for 2015:IV based on observing these past public signals. Specifically, we can construct models that use past tournament outcomes as predictors, and compare manager predictions to what our model says they "should" have predicted. We denote discrepancies between manager and model predictions as ex ante prediction errors. Under the assumption that tournament outcomes are the key (only) relevant signal, finding systematically optimistic ex ante prediction errors would be consistent with overconfidence bias.

A concern could be that it is too strong to assume that tournament outcomes are the only relevant signal, and that managers have access to some kind of additional, private signals. We check robustness to this issue in three ways.

First, we can use variation in how many signals, public and potentially private, that managers have observed, due to variation in experience. If managers observed private signals before starting the job, they could start with rationally overconfident priors, but as they gain experience and observe more tournament outcomes, predictions should come closer to our prediction models. Managers might instead observe private signals once the job starts, in each period of employment, but they should still learn from these, and thus become better at predicting future performance with

managers yields the (uniform) prior distribution (law of iterated expectations). As discussed in Section I, we did not pursue this avenue because of concerns about managers not understanding the relatively complex methods needed to incentivize belief distributions.

[21] The test uses the restriction that, for individuals with a given true type, the modal signal must be that they are that type (otherwise the signal is not informative). The test therefore checks whether, among those who predict a given quintile, the modal individual has zero ex post prediction error.

[22] The $p$-value of the test is close to 1. Burks et al. (2013) elicit the predictions of truckers about their modal quintile on tests of numerical ability and IQ, and reject the Bayesian model because the ex post errors they observe are quite pronounced; e.g., for a numeracy test, 95 percent of those in the worst quintile in terms of numerical ability predict being in quintile 3 or higher, and the average error is about 2.6 quintiles. By comparison, panel B of Figure 1 shows that among our managers who end up in the worst quintile for the fourth quarter, about 51 percent predict being in quintile 3 or higher, and the average error relative to ex post realizations is about 1.8 quintiles.

more experience, i.e., have smaller ex post prediction errors. We can check these predictions in the data.

Second, we can use our measure of memories of past signals, checking whether there is overly positive memory about past tournament outcomes, and whether this is related to optimistic ex ante prediction errors. Such findings would be consistent with managers being motivated to bias predictions toward overconfidence, but not a Bayesian explanation for manager optimism based on private signals.

Third, we employ a structural model to further discipline explanations based on private signals. In the model we use data on past tournament outcomes to calibrate what managers' priors should be based on public signals, and then ask if there is a structure of additional, private signals that can allow the model to come close to rationalizing manager predictions. The information about priors based on public signals places relatively stringent restrictions on the ability of private signals to rationalize manager predictions.[23]

Our approach of benchmarking manager predictions against prediction models makes sense only if past tournament outcomes are informative for predicting future performance; Table 1 shows that this is in fact the case. The table gives the frequencies of managers ending up in different tournament quintiles in quarter $t$ conditional on quintile in $t-1$. We denote this transition matrix $\hat{Z}$. The transition probabilities indicate that the quintile outcome in any given quarter $t-1$ is predictive of the quintile outcome in quarter $t$: the modal outcome is for that same quintile to occur in the next quarter.[24]

## B. *Testing for Overconfidence*

Our first step in testing for overconfidence bias is to identify the best performing prediction model out of a set of candidate prediction models. A natural class of models to consider given our data is panel regression models, which predict a manager's future performance based on lagged values of past performance. For a given model we use multinomial logit estimation to generate predicted probabilities of each quintile ranking in 2015:IV for each manager, and select the quintile with highest probability as the prediction. Within this class of models, two specific questions we consider are (i) what is the optimal number of lagged performance outcomes for maximizing predictive power; (ii) should performance outcomes in

[23] The restrictions are stringent because, given the informative public signals that they observe, many managers should have strong beliefs about what quintile they are in, if they are Bayesian. When such beliefs are strong, private signals must also be strong to move beliefs enough to change the mode, but then this limits the number of managers who can be overconfident (relative to underconfident). This is because, according to the law of iterated expectations, fixing the amount of underconfidence, there can either be many individuals weakly adjusting their beliefs in the direction of overconfidence, or there can be a few people strongly adjusting their beliefs in the direction of overconfidence, but there cannot be many people adjusting beliefs strongly in the direction of overconfidence. While our main analysis assumes that managers combine public signals with flat initial priors, we also consider whether overconfident initial priors (at the start of the job) could rationalize the overconfidence that we observe for experienced managers.

[24] Another takeaway is that quintiles 1 and 5 are particularly informative, in the sense of being persistent: if a manager is in one of these quintiles in $t-1$, the likelihood that they will be in the same quintile in $t$ is relatively high. This feature of the information structure could be consistent with a normal-shaped distribution of underlying manager ability: the mass of managers in the middle would have relatively similar abilities, choose similar effort levels, and thus have tournament outcomes that are largely random; managers in the tails would be quite different from everyone else in terms of ability and thus consistently have the worst or best outcomes.

Table 1—Quintile-to-Quintile Transition Matrix $\hat{Z}$

| | Fractions of managers | | | | |
|---|---|---|---|---|---|
| | Quintile in $t$ | | | | |
| Quintile in $t-1$: | 1 | 2 | 3 | 4 | 5 |
| 5 | 0.05 | 0.11 | 0.15 | 0.26 | 0.43 |
| 4 | 0.11 | 0.17 | 0.21 | 0.27 | 0.24 |
| 3 | 0.17 | 0.23 | 0.27 | 0.20 | 0.13 |
| 2 | 0.24 | 0.26 | 0.23 | 0.17 | 0.10 |
| 1 | 0.43 | 0.23 | 0.20 | 0.09 | 0.05 |
| Observations | 961 | 1,018 | 1,034 | 1,007 | 962 |

*Notes:* Best performance is quintile 5. The rows show the average proportions of managers achieving different quintile outcomes in the national tournament ranking for quarter $t$ conditional on a given quintile outcome in quarter $t-1$, using all quarters from 2008:I to 2015:IV. The number of observations differs across quintiles in $t-1$ due to attrition and opening of new stores.

a given past quarter be measured linearly in terms of percentile of performance, or nonparametrically with separate indicators for each quintile of performance?

It turns out that using a substantial number of lags (eight lags), and using the linear specification with percentile of performance for each past quarter, delivers the best predictive power in our model selection exercise. The exercise was based on cross validation, a simple machine learning technique that tests predictive power using randomly selected "hold-out" samples (for details see online Appendix F). The resulting prediction model can be written as

$$q_{i,t} = \alpha + \sum_{j=t-1}^{t-9} \beta_j y_{i,j} + \epsilon_{i,t}, \tag{1}$$

where the dependent variable $q_{i,t}$, is performance quintile for manager $i$ in quarter $t$, and independent variables are performance outcomes in earlier quarters, $y_{i,j}$, $j \in (t-1,\ldots,t-9)$. It is not surprising that the model does best when it includes a large number of lags, as this entails estimation on a sample of (relatively experienced) managers, for whom we have a large number of signals and thus better precision in assessing individual manager types.[25] The robustness checks include estimating models with fewer lags, and also using less parametric specifications.

Panel A of Figure 2 shows that the distribution of predicted quintiles from the regression model is slightly U-shaped, with the highest masses for quintiles 1 and 5.[26] Panel B shows that many managers made predictions that are substantially different from the predictions of the model, and these differences are much more frequent in the overconfident direction: 48 percent of managers bet on a higher quintile

[25] Better performance of the linear specification can be due to the fact that it provides a finer grained measure of performance, compared to the less-parametric but also coarser specification using quintile dummies.

[26] Given infinite signals, the distribution should converge to a uniform, but in a finite sample, there can be a U-shape because extreme outcomes are especially informative (Table 1). To see this, suppose there are 5 types of managers, and the worst and best types are quite likely to have an outcome of 1 or 5, respectively, and never get an outcome of 3. Suppose the remaining types have a more uniform probability of getting outcomes 2, 3, and 4, but also nonzero probabilities of getting 1 and 5. In a finite sample, due to chance, some intermediate types could have modes of 1 and 5, but no high or low types will have modes of 3.

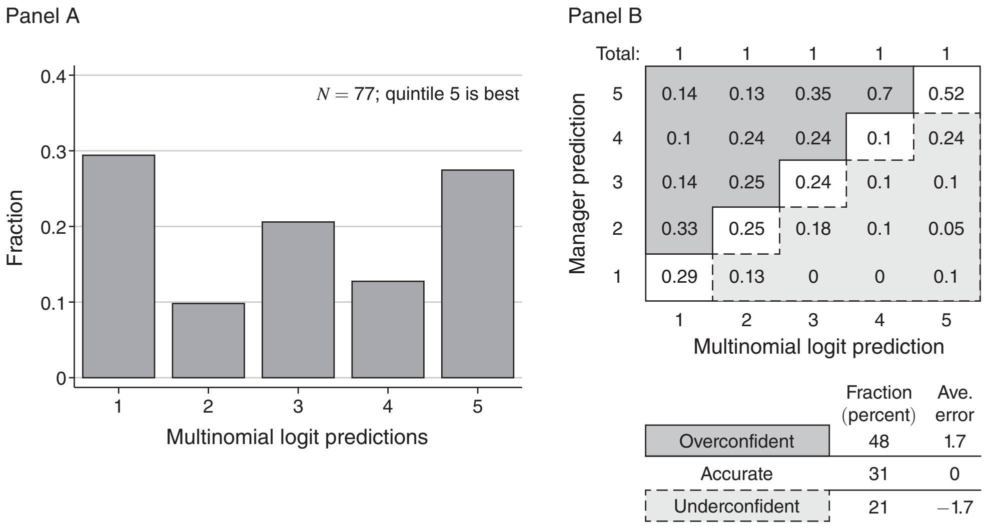


Figure 2. Distribution of Multinomial Logit Predictions and Comparison to Manager Predictions

*Notes:* Predictions are in terms of quintiles of fourth-quarter performance, with 5 being the best. Prediction errors are also in terms of quintiles.

than the model says was most likely for them, versus 21 percent betting on a worse quintile. The magnitude of the average error is substantial, 1.7 quintiles on average in both directions. The average ex ante prediction error is overly optimistic, involving a prediction that is 0.47 quintiles better than the model predicts. The estimated coefficients of the baseline model are reported in columns 1 to 4 of Table F2 in the online Appendix.[27]

The model parameters are estimated from a sample of tournament outcomes that have a random component, raising questions about statistical significance of the differences we find. Suppose that managers use the same model that we use, and are fully informed about the true parameters of the model; our model prediction could still differ from the manager prediction because it suffers from estimation error. We use bootstrapping to check whether the difference between manager and model predictions lies within the bounds of this error.[28] Specifically, we re-estimate the model 100 times, using samples drawn randomly from the data (with replacement), and each time generate predictions of the modal quintile for each manager. For a given bootstrap, we calculate the distance of each manager's bootstrapped prediction from the prediction of the model based on the original sample, using the Euclidean distance metric (results hold using alternative metrics or statistical tests, see online Appendix F.2).[29] Summing up these distances

[27] Most of the individual lag coefficients are not statistically significant individually, but this reflects correlated performance over time for managers. The coefficients are highly significant in a joint test ($\chi^2$; $p < 0.001$) and fit is improved by including all of the lags.

[28] In contrast to this scenario, if the managers have to estimate the parameters as we do, using the same data, then their predictions should accord with our predictions precisely, and we do not need confidence intervals to reject that the model and manager predictions are the same.

[29] The total Euclidean distance is just a monotonic transformation of the fraction of managers who differ from the model. We focus on Euclidean distance because it naturally generalizes to situations where we are computing

across all managers gives a total (Euclidian) distance between a given bootstrapped distribution of bets and the distribution of bets generated by the original model. This procedure yields a distribution of 100 distances, which gives bounds on the sensitivity of our model predictions to sampling error. The distance of observed manager predictions from the original model predictions lies far in the tail of the bootstrapped distances (beyond the ninety-ninth percentile); see Figure F1 in the online Appendix. Using a similar approach it is also possible to reject at the 1 percent level that the degree of asymmetry toward overconfidence that we observe, comparing manager predictions to model predictions, lies within the bounds of the asymmetry that could be generated by noise in our model.[30] Thus, estimation error in our model does not appear to explain why manager and model predictions differ.

A different reason why manager predictions might deviate from those of our selected model is if some of the assumptions underlying our model are wrong, and managers therefore use an alternative model. Table F1 in the online Appendix summarizes robustness checks based on a range of modifications to our multinomial panel regression model, with corresponding coefficient estimates reported in Tables F2 and F3.

For example, one robustness check addresses the fact that our candidate prediction models implicitly assume that managers come to the job with flat priors. If managers start the job with (potentially rational) overconfident priors, however, due to private signals received before starting the job, then they could make predictions that are more confident than our incorrectly specified model initially, although this should diminish with experience if they are Bayesian. We look at a model with only three lags, because such a model can include managers with as little as one year of experience, in contrast to the eight-lag model, which uses only managers with more than two years of experience. Table F1 shows a very similar degree of manager overconfidence relative to this model, so there is no sign that overconfidence is larger in a sample that includes managers who have had less feedback, contrary to an explanation based on private signals received before starting the job.[31] Additional analysis provided in Figure I2 in the online Appendix shows that the magnitudes of ex ante prediction errors are also not decreasing with experience, comparing managers with less than two years of experience to managers with at least two years of experience (we also find that ex post prediction errors do not diminish with experience).[32]

the distance between vectors that do not all have 0 or 1 entires, something that arises later in our analysis when we simulate some versions of our structural model.

[30] For each of the bootstraps we calculate the fraction of bootstrapped predictions that are overconfident relative to the original model minus the fraction of bootstrapped predictions that are underconfident. This yields a distribution of 100 differences. The corresponding difference comparing actual manager predictions to the predictions of the original model is beyond the ninety-ninth percentile of the distribution (see Figure F1 in the online Appendix).

[31] This is not to say that managers do not start the job with overconfident priors; indeed, if we look solely at recently hired managers (less than one year of experience) we see a distribution of predictions that is skewed toward higher quintiles, similar to what we observe for the sample as a whole (see Figure I1 in the online Appendix). In a robustness check we investigate whether tournament performance might become more variable and less informative with manager experience for some reason; if so this could be a Bayesian explanation for lack of learning. We find, however, little evidence for this (see online Appendix I).

[32] The rate of ex post accuracy of predictions is 35.71 percent for managers with less than two years of experience versus 35.16 for managers with two or more years. This argues against an explanation based on managers learning from private signals each period on the job. Note that we would miss some manager learning, if those who do learn their types tend to leave the firm, but this does not alter the fact that those managers who remain should be more accurate than inexperienced managers, if they are Bayesian. Furthermore, as discussed at the end of the paper, we find that manager overconfidence is not significantly related to the probability of remaining at the firm.

Other robustness checks summarized in Table F1 address implicit assumptions of our set of candidate prediction models regarding time stationarity of the environment, and time stationarity within managers, e.g., by using only recent quarters to estimate model parameters, or using only quarters from a manager's current store.[33] We consistently see that manager predictions are significantly more confident than the corresponding regression model predictions.

As another type of robustness check, we investigate whether manager predictions might be well explained by the use of some simpler, rule of thumb type predictors based on past tournament outcomes. If so, this might indicate bounded rationality, but motivated beliefs would not be needed to explain the data. One seemingly natural rule of thumb is the manager's most frequent quintile outcome in the past. Calculating each manager's modal quintile over past quarters, and dropping managers who do not have a unique mode, yields the distribution of historical modes. It turns out that manager predictions are substantially more confident than one would expect if they used the historical mode: 43 percent of managers predict a higher quintile for 2015:IV than their historical modal quintile, compared to 25 percent predicting a lower quintile, and the average prediction error is overly optimistic, by about 0.41 quintiles (see Figure H1 in the online Appendix). In additional robustness checks we consider alternative rules of thumb, and in all cases, manager predictions are significantly more confident than the corresponding rule of thumb prediction. These results are summarized in the online Appendix in Table H1.[34]

### C. *Testing for Biased Memory*

The evidence so far is consistent with managers exhibiting persistent overconfidence bias, under the assumption that past tournament outcomes are the key signals managers should be using to predict future tournament outcomes. Models of motivated beliefs offer an explanation for how individuals can be persistently overconfident, and also generate an additional testable prediction, that individuals may be motivated to have overly positive memories of past performance. In this section we investigate whether there is support for this prediction, using our measure of manager memory of rank in 2015:II.

Thus, there does not seem to be much scope for differential attrition on the basis of overconfidence to play a role in explaining our results.

[33] One source of nonstationarity in the environment could be patterns of manager turnover, which change the composition of manager abilities at the firm, and thus alter the predictiveness of lagged tournament outcomes over time. Besides checking robustness to estimating our regression model using only recent quarters, we check for time trends in the elements of the transition matrix $\hat{Z}$, and find little evidence of nonstationarity (see online Appendix G).

[34] A different potential confound would be if managers are inattentive in our survey, and use a heuristic of just choosing the top row of the choice table in our prediction measure. Because the top row corresponded to the best quintile, this could lead to the appearance that managers are systematically overconfident relative to our prediction models, when in fact it just reflects inattention. There are several reasons why this does not seem to drive the results. First, as shown in panel B of Figure 2, much of the overconfidence we find is not due to managers predicting the best quintile. Second, we used the same table format to elicit manager predictions about their performance quintile in one of our incentivized math tasks (see Part 5 of the instructions provided in online Appendix U). There we see only 11 percent predicting the best quintile, in contrast to 30 percent predicting the best quintile in the workplace tournament (panel A of Figure 1). This suggests that the table format per se does not lead to the extent of predicting the best quintile that we find for manager predictions about the workplace tournament. Finally, we did not elicit manager memories using a choice table, so a choose-the-first-row heuristic would not explain evidence of biased memory, or a correlation of biased memory with overconfident predictions.

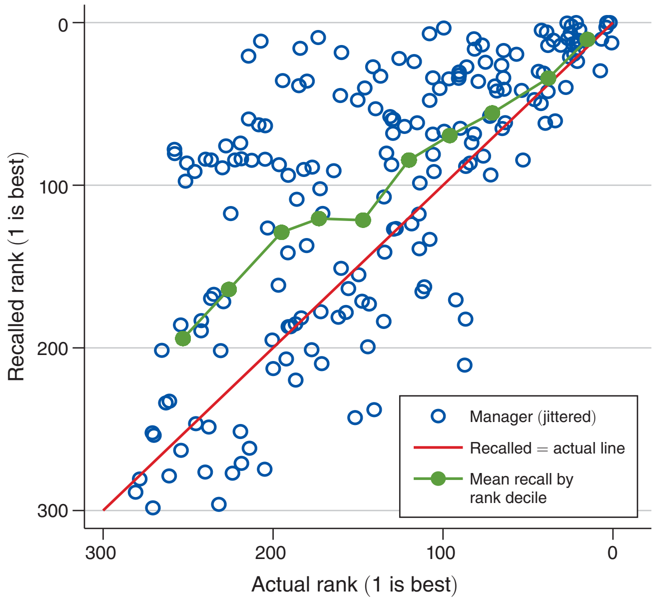


FIGURE 3. RECALLED PERFORMANCE FOR 2015:II, BY ACTUAL PERFORMANCE

Figure 3 displays the raw data from our elicitation of manager memories, about rank in 2015:II, with values jittered slightly to preserve manager confidentiality.[35] The $x$-axis measures individual managers' actual ranks in the second quarter, with 1 being the best, and the $y$-axis shows managers' recalled ranks. Whereas in the rest of the paper higher numbers indicate better performance, in this figure we use smaller numbers for better performance, since the question used to collect the recall data asked about rank.

A first observation about Figure 3 is that the best performing managers in 2015:II were quite accurate in their recollections. The lack of upward errors is mechanical, but managers also have only small errors in the downward direction. This shows that at least these managers could recall past rank accurately, and furthermore, it is consistent with a motivation to have positive memories; with such a motivation, recalling accurately that one had a top performance is attractive, whereas recalling lower than actual performance is counterproductive.

A second notable feature of Figure 3 is a clear increase in the frequency of managers with inaccurate memories, as soon as one goes below the top ranks, and a tendency for these memory errors to be asymmetric in the direction of recalling better than actual performance. The correlation between an indicator for being inaccurate, and rank in 2015:II, is statistically significant (Spearman; $\rho = -0.27$; $p < 0.001$). As shown in the figure, the tendency for recall errors to be in the better than actual direction results in the average recalled rank, by decile of actual rank, always being above average actual rank. Overall, 56 percent of managers have

[35] Jittering involves adding a small random mean zero perturbation to the values. Without jittering, the firm could in principle use its knowledge of the second-quarter ranking to infer individual managers' reported memories from Figure 3.

"flattering" recall errors, compared to 24 percent having "unflattering" errors, and the average recall error involves recalling a performance that is more than 30 ranks better than actual, with the error significantly different from zero ($t$-test; $p < 0.001$). The asymmetry in recall errors is strongly apparent for managers in the middle of the second-quarter performance distribution, so it is not driven by managers at the bottom of the distribution for whom floor effects force recall errors to be in the positive direction. This asymmetry matches the prediction of models of motivated beliefs in which individuals are motivated to have positive memories; as performance worsens, managers may want to recall a better than actual performance.

A third observation about Figure 3 is that average recalled rank (by decile of actual rank) does decline with actual rank, and the correlation of actual and recalled rank is substantial, 0.71 (Spearman; $p < 0.01$). Thus, manager recollections are neither completely random nor completely self-serving, but rather are tethered to actual past performance. This is consistent with manager memories being subject to some "reality constraints," as is typical in models of motivated beliefs. We also observe variation in the extent of memory distortion for a given actual rank, which could reflect some randomness inherent in the memory distortion technology or could indicate individual heterogeneity in manager costs or benefits of memory distortions.[36]

The conclusions from Figure 3 also hold up in regression analysis, which allows addressing some potential concerns related to the fact that performance in 2015:II is not randomly assigned. Column 1 of Table 2 presents results of a Probit regression where the dependent variable equals 1 if a manager has an inaccurate memory and 0 otherwise; the results show that a 1 standard deviation increase in the second-quarter performance is associated with a decrease of 0.12 in the probability of having inaccurate memory. Such a relationship could, however, be endogenous due to omitted variables. For example, some manager trait, e.g., lower cognitive ability, might foster both worse performance and inaccurate memory.[37] This suggests a benefit of controlling for manager ability.[38] Column 2 of Table 2 shows that being inaccurate is still significantly related to performance in 2015:II, controlling for manager ability by using performance in 2015:III as well as the mean performance across all pre-2015:II quarters. Thus, it is not good performance in general that is associated with accurate memory of 2015:II, but rather something special about a good performance in 2015:II. Controlling for some manager characteristics that could potentially affect both performance and memory—gender, age, and experience—leaves the results unchanged.[39] In terms of the direction of recall errors, columns 3 to 6 show that having a worse second-quarter performance is mainly associated with a

[36] Regarding potential sources of randomness in the technology, there might be idiosyncratic shocks to the arrival of the types of information that can be used to construct positive memories, leading to variation in memory distortion across managers in a given quarter (and across quarters for a given manager).

[37] This would be akin to the Dunning-Kruger effect, in which low ability people make worse predictions about relative performance (see Kruger and Dunning 1999), but for memory rather than predictions.

[38] A different explanation could be related to the convexity of the incentive scheme; managers who are typically in the worse quintiles of performance might perceive a relatively lower incentive to remember correctly, since the marginal benefit of effort is (locally) lower. This would also suggest controlling for manager ability.

[39] Interestingly, managers with more experience have a lower probability of having inaccurate memory, but the relationship is arguably relatively weak, as it takes about 3.7 years (one standard deviation) of additional experience for the probability to drop by 0.09. The fraction of managers who have inaccurate memories is 0.83 for managers with two years or less experience, compared to 0.78 for managers with more than two years of experience.

Table 2—Inaccurate Memory and Recall Errors as a Function of Actual Second-Quarter Performance

| | Inaccurate memory | | Flattering memory | | Unflattering memory | | Recalled-actual performance | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Performance percentile in 2015:II | −0.12 | −0.12 | −0.07 | −0.11 | −0.06 | −0.01 | | |
| | (0.03) | (0.04) | (0.04) | (0.05) | (0.03) | (0.04) | | |
| Inaccurate memory | | | | | | | 38.64 | 39.09 |
| | | | | | | | (6.07) | (7.30) |
| Performance percentile in 2015:III | | 0.02 | | 0.07 | | −0.06 | | 0.04 |
| | | (0.04) | | (0.05) | | (0.04) | | (5.24) |
| Mean performance percentile pre-2015:II | | −0.02 | | 0.00 | | −0.03 | | −2.43 |
| | | (0.03) | | (0.04) | | (0.03) | | (4.45) |
| Female | | 0.04 | | 0.12 | | −0.09 | | 11.64 |
| | | (0.06) | | (0.08) | | (0.07) | | (11.00) |
| Age | | 0.02 | | 0.01 | | 0.01 | | −0.24 |
| | | (0.03) | | (0.05) | | (0.04) | | (6.89) |
| Experience | | −0.08 | | −0.09 | | 0.01 | | −5.10 |
| | | (0.04) | | (0.05) | | (0.04) | | (7.27) |
| Constant | | | | | | | 0.11 | −5.33 |
| | | | | | | | (0.86) | (6.09) |
| Observations | 172 | 149 | 172 | 149 | 172 | 149 | 176 | 151 |
| Estimation method | Probit | Probit | Probit | Probit | Probit | Probit | OLS | OLS |
| Adjusted $R^2$ | | | | | | | 0.050 | 0.046 |
| Pseudo $R^2$ | 0.092 | 0.123 | 0.013 | 0.061 | 0.015 | 0.045 | | |

*Notes:* Columns 1 to 6 report marginal effects from probit regressions. Columns 7 to 8 report OLS estimates. The dependent variable for columns 1 and 2 is an indicator for a manager's recalled performance for 2015:II being different from their actual performance by $+/-$ 10 ranks (the elicitation gave an incentive to be accurate within this range). The dependent variables for columns 3 to 4, and 5 to 6, are indicators for remembering a better than actual performance in the second quarter by more than ten ranks, or worse than actual by more than ten ranks, respectively. The dependent variable for columns 7 to 8 is constructed by taking the difference between recalled rank and actual rank, and multiplying by $-1$, so that positive numbers indicate recalling a better than actual performance. Independent variables are standardized, so coefficients give the change in the dependent variable (level or probability) associated with a one standard deviation increase in the independent variable. Performance percentile independent variables are constructed as (recalled) rank expressed as a fraction of the worst rank in the corresponding quarter, and then reversed so that higher numbers reflect better performance. Robust standard errors are in parentheses.

higher propensity to have errors in the overly positive direction; there is a weaker relationship of performance to the propensity to have unflattering errors.[40] In columns 7 to 8 the dependent variable is the difference between recalled and actual performance, and the coefficient on an indicator for inaccurate memory shows that the average recall error is significantly different from zero in the direction of recalling better than actual performance.

Robustness checks, reported in online Appendix J, add controls for additional factors that might conceivably affect the probability of misremembering, or the particular performance that is remembered. These include the degree to which second-quarter performance deviates from a manager's typical (mean or median) pre-second-quarter performance, in case this affects memorability, and the variance of manager past performance, in case managers with more variable performances are less likely to remember a given quarter's performance. We also control for the

[40] The imprecision in the estimates means that the difference in coefficients across columns 4 and 6, for second-quarter performance, is not statistically significant ($p < 0.36$).

time elapsed between end of the second quarter and when memory was elicited, to see if shorter duration is associated with more accuracy, and for a proxy for being motivated by the incentives we offer for memory accuracy: willingness to work on an incentivized addition task. Other controls include proxies for traits of attentiveness and cognitive ability, based on incentivized questions testing knowledge and understanding of details of the firm's incentive scheme.[41] Further controls include self-reported manager traits, and other summary statistics of past performance, such as maximum, median, and minimum career performances. Performance in the second quarter remains the key explanatory variable for memory of the second quarter, while these additional factors are by and large not significantly related to manager memories.[42]

Finally, in additional analysis we verify more rigorously that the asymmetric nature of recall errors about rank is not driven by floor effects, and we check whether biased memory is also present for other performance metrics besides rank. In online Appendix J we report results of estimating the specifications in column 7 and 8 of Table 2 but excluding managers in the bottom quintile of second-quarter performance, and find similar results. This goes against an explanation based on floor effects (Table J3 in the online Appendix). In online Appendix K we analyze some additional memory measures included in the lab-in-the-field study, which asked managers to remember some of the different submetrics that determined a manager's rank in 2015:II. We also find a systematic tendency toward overly positive memories for all of the submetrics, pointing to the robustness of the tendency to have overly positive memories of past performance on a range of metrics.[43] Taken together, our findings are consistent with a specific structure of recall errors predicted by a motivated beliefs explanation.

### D. *Testing for a Link between Biased Memory and Overconfident Predictions*

We have seen that, in the aggregate, managers are both more confident about future rank than seems justified based on past histories, and also overly positive in their memories about past rank. Models of motivated beliefs predict, however, that these should be positively correlated at the individual level. This is what we investigate in this section.

Table 3 presents regressions that investigate the hypothesized link between biased memory and manager predictions. In column 1 the dependent variable is manager predictions about the most likely quintile in 2015:IV. The estimation method is interval regression, which models the conditional mean of manager predictions

[41] One set of questions asked managers if they knew the maximum and minimum values used by the firm to score relative performance on the four dimensions, and another question tested understanding of the implications of the multiplicative value of the scheme, namely that higher variance in performance across the four dimensions yields a lower bonus, holding constant average performance.

[42] One exception is manager experience: managers who are quite experienced have a tendency to recall worse performances, leading to a modest decrease in the proportion with overly positive memories, and increases in both the proportion with accurate memories, and the proportion with overly negative memories. We discuss a possible interpretation of this pattern in the online Appendix.

[43] Interestingly, the rate of memory accuracy is highest, and the asymmetry in errors is least pronounced, for area bonus, a performance metric that is more group based; this is suggestive of another comparative static of some motivated beliefs models, in which the reason to distort memory comes from a desire to have positive beliefs about the self, as opposed to about outcomes in general (general optimism).

Table 3—Manager Predictions and Overconfidence as a Function of Recalled Second-Quarter Performance

| | Manager prediction | | Overconfident (rel. to multi. logit) | | Overconfident (rel. to historical mode) | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Recalled performance quintile for 2015:II | 0.55 | 0.43 | | | | |
| | (0.17) | (0.17) | | | | |
| Flattering memory about 2015:II | | | 0.20 | 0.20 | 0.18 | 0.15 |
| | | | (0.10) | (0.10) | (0.08) | (0.08) |
| Performance percentile in 2015:II | 0.41 | 0.21 | −0.14 | −0.14 | −0.07 | 0.01 |
| | (0.18) | (0.17) | (0.05) | (0.06) | (0.04) | (0.05) |
| Performance percentile in 2015:III | | 0.62 | | 0.00 | | 0.06 |
| | | (0.15) | | (0.06) | | (0.04) |
| Mean performance percentile pre-2015:II | | 0.06 | | −0.11 | | −0.19 |
| | | (0.10) | | (0.06) | | (0.03) |
| Female | | −0.15 | | −0.14 | | −0.04 |
| | | (0.23) | | (0.11) | | (0.08) |
| Age | | −0.09 | | −0.00 | | −0.07 |
| | | (0.12) | | (0.07) | | (0.05) |
| Experience | | 0.02 | | −0.06 | | −0.01 |
| | | (0.14) | | (0.08) | | (0.05) |
| Constant | 3.08 | 3.16 | | | | |
| | (0.11) | (0.19) | | | | |
| Observations | 170 | 148 | 75 | 75 | 128 | 120 |
| Estimation method | Int. reg. | Int. reg. | Probit | Probit | Probit | Probit |
| Pseudo $R^2$ | 0.101 | 0.157 | 0.115 | 0.152 | 0.044 | 0.187 |

*Notes:* Columns 1 and 2 report marginal effects from interval regressions, which correct for the interval nature of the dependent variable (right and left censoring for each interval); the dependent variable is the manager's prediction about fourth-quarter performance quintile. Columns 3 to 6 report marginal effects from probit regressions. The dependent variable for columns 3 and 4 is an indicator for whether a manager predicted a higher quintile than the quintile predicted by the baseline (eight lag) multinomial logit model. The dependent variable for columns 5 and 6 is an indicator for whether a manager predicted a higher quintile than their historical modal quintile. Independent variables are standardized, so coefficients give the change in the dependent variable (level or probability) associated with a one standard deviation increase in the independent variable. Performance percentile independent variables are constructed as (recalled) rank expressed as a fraction of the worst rank in the corresponding quarter, and then reversed so that higher numbers reflect better performance. Robust standard errors are in parentheses.

while accounting for the fact that the dependent variable is measured in intervals (right and left censored).[44] Independent variables are standardized. Column 1 shows a significant positive relationship between recalled performance from the second quarter and predictions about the fourth quarter, controlling for actual second-quarter performance. Column 2 adds more controls for past performance, and manager traits. The coefficient on recalled performance remains significant, and implies that a one standard deviation increase in recalled second-quarter performance is associated with predicting about 0.5 quintiles higher performance in the fourth quarter. The coefficient on actual second-quarter performance is half the size, and not statistically significant, consistent with managers basing predictions mainly on remembered rather than actual past performance.

[44] We prefer interval regression over multinomial logit in this case as the goal is to model the conditional mean rather than produce predictions of modal quintiles. Results are similar, however, using multinomial logit: more self-flattering memories of the second quarter are associated with a lower probability of predicting low quintiles.

Columns 3 to 6 of Table 3 use two different indicators for overconfidence as the dependent variables, to check whether having overly positive memories is associated with overconfidence about the future. There is a significantly higher probability of being overconfident, according to both indicators, if a manager has a flattering (overly positive) memory of the second quarter. The coefficient on manager experience is small and not statistically significant, so the likelihood of overconfidence does not diminish with experience. In case tenure at the firm is endogenous to overconfidence, we checked robustness to excluding experience from the regression, but other coefficients are qualitatively unchanged (at the end of the paper we discuss empirical evidence that suggests tenure is not in fact significantly affected by overconfidence).

Online Appendix L presents robustness checks on whether these results extend to using other types of indicators for overconfidence, to using indicators of underconfidence, and using nonbinary measures of manager prediction errors and recall errors. Across a wide range of different models, overconfident predictions are associated with overly positive memories.[45]

Other robustness checks, reported in Table L7 of the online Appendix, show that results are similar if we add additional controls. These include other moments of the distribution of past performance (median, mode, maximum, minimum), in case manager memories of the second quarter are correlated with these other summary statistics of past performance. The regressions also include days between eliciting manager predictions and the end of 2015:IV, as a possible determinant of prediction accuracy; an indicator for valuing the magnitude of incentives offered in our study, as proxied by willingness to solve incentivized addition problems; proxies for traits of attentiveness and cognitive ability, based on incentivized measures of knowledge and understanding of details of the firm's incentive scheme, in case biased memory and overconfidence might be related due to an omitted variable of low cognitive ability; and controls for other manager traits. These controls are generally not significantly related to managers' overconfident predictions, and leave the key result unchanged, that overconfidence is significantly related to biased memory. Finally, Table L8 in the online Appendix shows robustness checks on whether the relationship of overly positive memories to overconfidence remains similarly strong as manager experience increases; in our various specifications, interaction terms between the indicator for overly positive memories and manager experience are not statistically significant, but the point estimates suggest that if anything the relationship is getting stronger with experience. In summary, our reduced form analysis finds support for a signature prediction of the motivated beliefs explanation for overconfidence, that persistent overconfidence about the future goes hand in hand with overly positive memories of past feedback.

## III. Structural Analysis

We complement the reduced form analysis with estimation of a structural model. This provides a way to discipline, and evaluate in quantitative terms, some explanations for the data that are not fully addressed by the reduced form analysis.

[45] Focusing on the 42 regression specifications that include the full set of controls, 41 have a coefficient for the measure of manager recall that is of the expected sign, and 30 are statistically significant.

As a baseline we start with a simple model of Bayesian learning from public signals. This enables us to check whether our reduced form results on overconfidence are robust to using an explicit Bayesian benchmark. It also allows us to refine some of our reduced form tests for confounds, e.g., whether overconfidence could be explained by private signals received before starting the job. The reduced form analysis tests the qualitative prediction that such overconfident priors should be corrected with experience using, e.g., an arbitrary threshold of seemingly substantial experience, two years. Our estimated structural model can go further, generating quantitative predictions about how quickly Bayesians should learn in our setting, and whether two years is a long enough time horizon to correct overconfident priors.

We next extend the model in two ways. Each extension incorporates restrictions that can be easily stated within the Bayesian framework, but which are difficult to translate into a reduced form approach.

The first extension incorporates private signals to further discipline explanations based on learning from such signals. One reason this is useful is because there are signal structures that are not addressed by our reduced form approach of looking at how prediction errors relate to experience. For example, if a private signal about a transitory shock affecting store performance is received right before we elicit manager predictions, this is not something that a manager could have learned with previous experience, but it could explain a deviation from our reduced form prediction model that uses only public signals. In the context of our structural model, which incorporates the priors managers should have based on past public signals, we can let the data tell us what structure of additional, private signals would bring the model as close as possible to manager predictions, including signals received at the time that we elicit manager predictions. We can assess whether this signal structure has plausible features, and whether this best fitting model can come close to the data in quantitative terms, rationalizing manager predictions within standard confidence intervals.

The second extension explicitly models a process of belief formation based on biased memory, along with sophistication and naïveté about such distortions, to see if this can help explain the data in quantitative terms. Our reduced form analysis on memory, by contrast, was qualitative, establishing support for a directional prediction, that overly positive memory should be associated with overconfident predictions, but not asking whether biased memory can explain the prevalence or size of overconfidence that we observe. Our structural model generates for each manager a prediction about the presence and extent of overconfidence, and we can assess how close the model comes, in quantitative terms, to matching the data on manager predictions. The rest of this section briefly discusses how we formulate the baseline Bayesian model and then turns to our two extensions.

### A. *Baseline Structural Model of Bayesian Prediction*

The baseline model assumes that there is a finite number of periods $t = 1, 2, \ldots, T$ corresponding to quarters. Each manager $k$ has a type $a_k$ that takes on a fixed value between 1 and 5 and is time invariant. Every period a public signal $s_{k,t}$ is generated for each manager, taking on an integer value between 1 and 5. This is manager $k$'s quintile in the quarterly tournament in period $t$. A manager's signal is a stochastic function of the manager's type $a_k$, i.e., $s_{k,t}$ depends partly on type but partly on luck. Denote by $p_t(s|a)$ the probability of a given signal $s$, conditional on a particular type $a$, in time period $t$. All information about the probabilities of signals associated with different types can then be summarized in a five-by-five "type-to-signal" matrix denoted $P_t$. Each row of the matrix corresponds to a type, and moving across the columns the $p_t(s|a)$'s give the probabilities of observing different signals for that type.

At any given time a manager will have a belief distribution $f$ that captures the probabilities that the manager assigns to being each of the possible types, with $f_{k,t}(a)$ denoting the belief that individual $k$ is of type $a$ in time period $t$. Beliefs about types also give rise to beliefs about what signal will be generated at the end of period $t$. Manager posterior beliefs about signal probabilities are denoted $g$, with $g_{k,t}(s) = \sum_a f_k(a) p_t(s|a)$. For example, if a manager thinks there is a 50/50 chance of being type 5 or type 4, then $g_{k,t}(s)$ is constructed by combining the probability distributions for rows 5 and 4 of $P$ with equal weights.

The goal is to establish, in the context of the model, what individuals should have believed about their probabilities of observing different signals, captured by $g_{k,t}(s)$. Given $g_{k,t}(s)$, it is possible to specify on which signal an individual should have bet (given our assumption the manager bets on the modal signal). As researchers we do not, however, observe manager type $a_k$, $P$, $f_{k,t}$, nor $g_{k,t}(s)$. Thus, these need to be estimated. For details see online Appendix M.1, which we briefly summarize here.

Estimation of the model is done in three steps. First we estimate a $\hat{P}$ using observable signal-to-signal matrices denoted $Z_t$ (the average of these is the transition matrix $\hat{Z}$ discussed above in Section II). Second, we start with uniform priors about each manager's type, and then use $\hat{P}$, each manager's history of tournament outcomes, and Bayes' rule to calculate a posterior distribution across types for each manager. Third, we use the posterior distribution across types to construct the posterior distribution of the probabilities of different signals, our estimate of $g_{k,t}(s)$, and identify each manager's modal quintile signal for 2015:IV. Then, as discussed, we suppose that managers bet on the signal they believe is most likely to occur.

We find that 45 percent of managers are overconfident relative to the baseline Bayesian model, compared to only 26 percent underconfident, an asymmetry that is quite similar to our reduced form results, and the average error is overly optimistic by about 0.4 quintiles. The prevalence and size of overconfidence is very similar among managers with more than two years of experience (47 percent overconfident). The model parameters imply that managers should learn relatively quickly, and two years is a long enough time horizon to correct even relatively extreme overconfidence in priors, supporting our reduced form approach of using a two-year cutoff. We find similar results when we check robustness to altering various assumptions in the baseline model, such as allowing for various forms of nonstationarity, or allowing for random choice errors (of plausible magnitude) in manager betting behavior. See online Appendices M.2 and N for details. Bootstrapping the model, we find that we can reject statistically that the model matches manager predictions. The Euclidian distance of manger predictions from the model, which is 210, lies far in the tail of the distribution of distances derived from bootstrapping the model. The difference in the observed fraction of managers who are overconfident versus underconfident relative to the model also lies far in the tail of the distribution of differences based on the bootstraps of the model (see online Appendices M.3 and M.4).

*Model Augmented with Private Information*.—We build on the Bayesian framework to allow for managers receiving additional, private signals, and let the data tell us what form of signal structure gives the model the best fit, and whether the model can come close to rationalizing manager predictions (details are in online Appendix P).

Suppose that after observing all public signals and having a posterior belief vector over types, managers receive a private signal. Since for Bayesians, signals are exchangeable in order, we can suppose that the private signal occurred in the last period, i.e., 2015:III, without loss of generality. These signals can either be interpreted as summarizing a sequence of signals drawn over time about an underlying type, or as a one-time set of signals, received right before the manager makes the prediction, that give information about a shock (potentially transitory) that will affect the manager's type starting in the next period. There are five potential private signals, 1 to 5, and the probability distribution over these signals may vary by manager type.[46] This private information can be summarized in a five-by-five type-to-signal matrix, which we denote $Q$, with the same interpretation as $P$ in the baseline model, i.e., each row corresponds to a type, and the entries give the probabilities of that type receiving the different possible signals. It is possible to estimate the elements of $Q$ that minimize the distance of the model predictions from the observed manager predictions. Since realizations of private signals are not observed, and cannot be fed into the model to generate predictions, the model predictions are based on calculating the expectation across different possible private signals conditional on type as well as the different possible manager types. The estimated $Q$ thus minimizes the distance between what managers actually bet and what the model predicts managers should bet on average (across many draws from the private signal distribution).[47]

The estimated best-fitting $Q$ is shown in online Appendix Table P1. While this $Q$ gives the best fit, this does not necessarily mean that model predictions come close to matching manager predictions. As explained intuitively in Section II, this is because the public signals already give information about managers' nonuniform priors, which places limitations on the ability of additional private signals to change a lot of managers' beliefs toward predicting higher quintiles. Indeed, we see that the model falls well short of generating the deviations from baseline model predictions that we see for managers. In fact, private information provides only a slight improvement over the baseline Bayesian model: 43 percent

[46] The proof of Theorem 4 of Benoît and Dubra (2011) shows that, when considering quintiles, considering at most five signals is sufficient to achieve the maximal distortion of beliefs toward overconfidence with private information.

[47] An advantage of a structural approach to addressing the issue of private signals is that it is straightforward to embed the relevant object—our matrix $Q$, as well as its restrictions (i.e., that the rows of $Q$ must sum to one, or in other words the average posterior equals the prior)—into the estimation procedure. We then find the best fitting $Q$ via a simulated methods of moments. In a reduced form model, by contrast, we would need to simulate both the unobserved signals generated by private information as well as the regression coefficient assigned to the impact of private information. Moreover, there is no simple way to imbed the restriction that the average posterior must equal the prior. Imbedding this restriction essentially amounts to moving to a structural approach, but without the explicit transparency provided by the structural model and $Q$.

of observed bets are overconfident, and 25 percent underconfident, relative to the predictions of the model with private information (recall this was 45 percent and 26 percent for the baseline model without private information).[48]

To assess statistical significance of the difference between the predictions of the model with private signals and manager predictions we simulate the model 100 times (see online Appendix P for details). For each simulation we calculate the Euclidean distance of the simulated bets from the average betting behavior. This yields a distribution of 100 distances. It turns out that the distance of manager predictions from the average betting behavior, which is 200, lies far in the tail of the simulated distances and we can reject that the manager predictions lie within the bounds of the randomness in the model at the 1 percent level. Also, the observed difference in the fraction of overconfident versus underconfident predictions lies far in the tail of the distribution of simulated differences, and we reject that the model can explain the observed asymmetry at the 1 percent level.

### B. *Model Augmented with Biased Memory*

In this section we augment the structural model to take into account the data on biased memory. This may be expected to help the model better match the data on manager predictions, for two reasons. First, if managers form predictions based on overly positive memories of past signals, this could help generate predictions that are more confident than the baseline structural model. Second, incorporating individual-level variation in memory distortion may help explain heterogeneity in manager overconfidence.[49] Our model incorporates individual heterogeneity in memory distortion in several ways, which are disciplined by the memory data and suggested by theory. In estimating the model we do not distinguish between different possible motivations for distorting memory, but merely seek to establish whether incorporating memory distortions can help better match observed manager predictions. More details on our approach are provided in online Appendix Q.

Our first source of heterogeneity is in terms of whether managers are motivated to misremember the past signals they have received. Some managers can be "unmotivated" and simply remember their past signals accurately. However, we also allow for the possibility that some managers are "motivated" to distort memories. We incorporate a technology for memory distortion by adding a "memory matrix," denoted $M$, to the baseline model. Each row corresponds to having received one of the public signals 1 to 5. Each column gives probabilities that the manager remembers signals 1 to 5 (i.e., quintile ranking). The data on manager recall provide a way to calibrate $M$. For any given quintile of actual performance, the matrix uses the empirical frequencies of remembering different ranks that fall in quintiles 1 to 5 (we report $M$ in online Appendix Q). The observed frequencies have several notable features. Memory distortions will most frequently occur in the overly positive direction; memories will

[48] For each manager the model generates a probability distribution over bets on the different quintiles of the performance distribution. To provide these summary statistics on the fractions overconfident, accurate, and underconfident, we use modal bet predicted by the model for each manager, and compare to the manager's observed bet.

[49] Indeed, we have seen that having overly positive memories is predictive of overconfidence relative to reduced form predictors (results are also similar measuring overconfidence relative to the baseline structural model, see Table Q1).

be correlated with actual signals; there will be variation in the extent of memory distortion conditional on a given signal. This latter feature introduces a second source of heterogeneity, within managers over time, which can be thought of as reflecting an inherent randomness in the memory technology.[50] Managers are assumed to update beliefs each time they receive a new signal but using the remembered signal rather than the actual signal (the remembered signal could be the same as the actual signal).

In formulating the model it is necessary to make assumptions about self-awareness, as discussed in Bénabou and Tirole (2002) (which they refer to as metacognition). One possibility is sophistication, in which case individuals do not have access to the actual past signal, but take into account the motives of past selves, and $M$, when updating beliefs. Even with full sophistication, the individual can still make overconfident predictions.[51] At the other extreme, one could assume motivated managers are completely naïve, treating remembered signals as the actual signals, also leading to overconfidence. Perhaps more realistically, there could be heterogeneity in the degree of self-awareness among motivated managers. We also allow for this third type of heterogeneity in the model and empirically estimate it, assuming specifically that managers are either fully sophisticated or fully naive, and then estimating how many managers are best matched by either assumption.[52]

Since little is known about the prevalence of motivated versus unmotivated individuals, or about different levels of self-awareness about memory distortion, we let the data inform us about the most appropriate assumption for each manager—sophisticated, naïve, or "unmotivated" (always remember signals accurately). To do this we run 100 simulations for each manager, under each of the three different assumptions. Taking the average betting behavior for each assumption, we assign the manager to the type that has the smallest difference between the average betting behavior and the manager's actual bet. We arrive at 40 percent naïfs, 31 percent sophisticates and 29 percent unmotivated.[53]

Our next step is to assess the ability of the model to generate overconfidence, and to fit the data on actual manager predictions. We do 100 simulations of the model with each individual fixed to their assigned type (no memory distortion,

[50] For more on this randomness, see discussion in footnotes for Section IIC.

[51] In the case where $M$ is fixed across individuals and over time and somewhat informative, as in our approach, Compte and Postlewaite (2004) note that given sufficient signals sophisticates should learn their true type. However, when $M$ can be prior or history dependent (see, e.g., Gottlieb 2010, 2014), beliefs may not converge to the truth, as $M$ may become uninformative. We do not have enough data to estimate different $M$ for different histories of signals, so our model does not capture this latter possibility.

[52] An advantage of a structural model is that it is immediately obvious how a sophisticate should update in response to a signal (since they are a Bayesian, other than the fact they happen to misremember past signals). In contrast, in a reduced form model, it is less clear how to treat the distribution over actual signals generated by a remembered signal. For example, the natural approach would be to run the regression for each possible path of actual signals implied by the remembered signals, and then take the weighted average. But this implies that an individual who thinks with 50 percent chance they had a signal of 2, and with 50 percent chance they had a signal of 4, would be treated similarly to an individual who thinks they had a 100 percent chance of a signal of 3. However, it seems natural that an individual would infer very differently in these two circumstances. If signals are highly informative, in the first case they could have a bimodal distribution with most weight on 2 and 4, while in the latter case, they could have a unimodal distribution with most weight on 3.

[53] One caveat is that turnover in the manager population could cause these sample estimates to be biased relative to the fractions present in the worker population as a whole. Suppose that managers who are sophisticated, or who are unmotivated to distort memory, are more likely to leave the firm over time, because those with low ability recognize this and leave the firm. In this case the sample that we use, which requires managers to be present long enough to have an estimated type, is missing some of the sophisticates and unmotivated managers who are present in the population as a whole.

naïve or sophisticated). This yields a bet for each manager for each simulation. Taking the average across the 100 simulations gives expected betting behavior for each manager. We find that the model generates average betting behavior that is substantially overconfident relative to the baseline structural model: 33 percent of managers are overconfident and 17 percent are underconfident. Recall that manager predictions entail that 44 percent overconfident and 25 percent underconfident relative to the baseline model.[54] Thus, the model with biased memory generates a similar difference of overconfident versus underconfident managers as the data on manager predictions, although the prevalence of both biases is still somewhat smaller than in reality. Moreover, comparing managers' observed bets to the predictions of our memory model, we see that deviations are largely symmetric: 24 percent of individuals are overconfident and 23 percent are underconfident relative to the memory model.

To assess whether the model's predictions are significantly different from manager predictions, we use the fact that the 100 simulations yield a distribution of 100 Euclidean distances from average betting behavior. The Euclidean distance of manager predictions from average betting behavior (conditional on our memory-augmented model being true) is 135, which lies at the ninetieth percentile of the simulated distances. The difference in the fractions of overconfident versus underconfident managers lies at the eighty-seventh percentile of the simulated distribution of differences. Thus, unlike for previous versions of our structural model, we cannot reject that this version matches manager predictions at conventional significance levels. Furthermore, the distance between manager predictions and the model, at 135, is substantially smaller than for other models, e.g., the distances for the baseline model and the model with private information are both at least 200. One concern is that the model comes closer to the data because the various sources of heterogeneity give extra degrees of freedom, but even with zero heterogeneity and assuming the type that gives the worst fit, e.g., assuming all motivated sophisticates, the distances between the model and manager predictions are still smaller than for other versions of the structural model (online Appendix Q). The model is clearly still far from perfect in terms of capturing all nuances, but we conclude that incorporating biased memory is a move in the right direction in terms of helping to explain the observed overconfidence in quantitative terms.

## IV. Discussion and Implications

The findings in this paper are consistent with managers being overconfident about their future relative performance in the workplace, despite substantial feedback. The evidence of overly positive memories of past feedback and a link between these and overconfident predictions point to an explanation based on motivated beliefs. This is not to say that motivated beliefs are the entire explanation for the observed overconfidence; there could be other factors at work as well, both rational and psychological.

Evidence of motivated beliefs and biased memory in the field has important implications for economic theory. It implies that overconfidence can be a persistent

[54] For each manager the model generates a probability distribution over bets on the different quintiles of the performance distribution. To provide theses summary statistics on the fractions overconfident, accurate, and underconfident, we use modal bet predicted by the model for each manager, and compare to the manager's observed bet.

phenomenon in field settings with feedback, in contrast to standard models of belief formation. It also changes the ways that individuals respond to feedback, relative to standard theories of information provision and optimal feedback, and it implies that variables that should not matter for behavior in standard models may influence decisions. For example, presenting feedback in ways that are less "ego threatening" might matter for belief updating. There are also implications for theories of optimal incentive design if agents are persistently overconfident.

A motivated beliefs explanation for overconfidence also has different implications for welfare and policy, compared to if overconfidence is a cognitive mistake. In particular, it becomes less obvious that one should implement policies to minimize overconfidence. For individuals, welfare losses that arise because of making choices based on biased beliefs could be offset by an intrinsic utility benefit of positive beliefs, or by benefits in terms of counteracting other biases. From the perspective of a principal, biased beliefs might lead managers to make mistakes on the job, but there could also potentially be offsetting benefits, e.g., if greater confidence counteracts self-control problems.[55] On the extensive margin, overconfidence might make managers overestimate the value of employment relative to the outside option, with the benefit to the principal of relaxing the participation constraint.

Although opening the black box of managerial performance is not the focus of this paper, our data can shed some light on whether and how manager beliefs feed into the ways that managers perform and make decisions. One caveat is that the sample of managers is relatively small, to study determinants of managerial performance, and there are limited outcomes on decision-making that we can study. Another caveat is actually a methodological implication of our evidence that beliefs are motivated. Once overconfidence is motivated it is endogenous, which may complicate efforts to understand the impact of overconfidence on outcomes such as performance. For example, suppose that some individuals have self-control problems in the form of present-biased preferences. Those with self-control problems may anticipate poor performance in the future, and thus implement overconfident beliefs. If the overconfidence does not completely counteract self-control problems, there could actually be a negative correlation between greater confidence and performance, but this would conceal a positive effect, because the counterfactual would have been even worse performance. This methodological implication is potentially important for interpreting past and future empirical research on overconfidence.

With these caveats in mind, we regressed different aspects of future manager performance (from 2016:I and 2016:II) on manager predictions about 2015:IV, as well as various binary indicators for overconfidence. It turns out that managers who are overconfident about 2015:VI do not do any worse, or better, in terms of overall future performance compared to other managers.[56] Digging deeper into the underlying dimensions of performance, however, there are differences.

[55] See, e.g., Hvide (2002); Bénabou and Tirole (2003); Fang and Moscarini (2005); Gervais and Goldstein (2007); Santos-Pinto (2008, 2010); de la Rosa (2011); and Foschi and Santos-Pinto (2017) for discussions of implications of biased beliefs for contract form, performance, and welfare.

[56] One confound is if overconfident managers are assigned to systematically different types of stores, which have characteristics that matter for future performance. For this reason the analysis controls for store characteristics. We also explore whether our various indicators for manager overconfidence, and measures of other manager traits, are significantly correlated with store characteristics, but find little evidence of a systematic relationship.

Overconfident managers have higher profits, but they also have worse customer service scores (these results are generally statistically significant but not in all specifications; online Appendix R provides details). The findings are intriguing, as they suggest the possibility that overconfidence might be associated with strengths and weaknesses on different aspects of the job. Overconfidence might also be related to a manager's tendency to stay at the firm, if it causes managers to value the job more relative to outside options. Using different indicators for overconfidence, point estimates suggest lower hazard rates of leaving the firm for overconfident managers, but these differences are relatively small and not statistically significant (see online Appendix Figure S1). One explanation for the weak relationship could be that the managers' overconfidence is not entirely job specific, and inflates estimates of their outside options as well.

To explore how manager beliefs are related to managerial decision-making, we related some indicators of management style to our measures of manager overconfidence. One finding is that overconfident managers tended to hire fewer assistant managers than recommended by store-specific guidelines provided by the firm (results are less precisely estimated for some of the binary indicators; details are in online Appendix T). This suggests a type of overconfidence in terms of being able to manage the store without additional help. It could also potentially contribute to higher profits, because hiring fewer assistant managers reduces the wage bill, but it could seemingly also have downsides, e.g., possibly harming customer service. Managers with overconfident predictions also exhibited a type of overconfidence in management style in a measure collected in the lab-in-the-field study. Specifically, overconfident managers were more likely to be willing to pay a cost, to be able to choose for a worker which of two brain teaser problems to try to solve, rather than allowing the worker to choose which one to solve (empirically, the two questions were equally difficult). Manager payoffs depended on the worker getting the answer correct; for more details see online Appendix T. This suggests that overconfident managers may think that they are better able to assess task difficulty for a worker than the worker himself, even in situations where this is unlikely to be the case. Although correlational and exploratory, these findings provide a starting point for future research on how overconfidence shapes managerial style.

A final point is that our analysis, like much of the literature, focuses on overconfidence. There is a smaller literature, however, that discusses underconfidence (e.g., the original Kruger and Dunning 1999 analysis finds that very competent individuals are underconfident). Our results demonstrate that while overconfidence is more prevalent in our field setting, some individuals exhibit underconfidence. This heterogeneity, and the reasons for it, are an interesting area for future research.

## REFERENCES

**Anonymous Firm.** 2016. "Historical Performance Data on Managers." Unpublished data. Accessed August, 2016.

**Barber, Brad M., and Terrance Odean.** 2001. "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment." *Quarterly Journal of Economics* 116 (1): 261–92.

**Barron, Kai.** 2021. "Belief Updating: Does the 'Good-News, Bad-News' Asymmetry Extend to Purely Financial Domains?" *Experimental Economics* 24 (1): 31–58.

**Bénabou, Roland, and Jean Tirole.** 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics* 117 (3): 871–915.

**Bénabou, Roland, and Jean Tirole.** 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies* 70 (3): 489–520.

**Bénabou, Roland, and Jean Tirole.** 2004. "Willpower and Personal Rules." *Journal of Political Economy* 112 (4): 848–86.

**Bénabou, Roland, and Jean Tirole.** 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2): 805–55.

**Bénabou, Roland, and Jean Tirole.** 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives* 30 (3): 141–64.

**Benoît, Jean-Pierre, and Juan Dubra.** 2011. "Apparent Overconfidence." *Econometrica* 79 (5): 1591–1625.

**Benoît, Jean-Pierre, Juan Dubra, and Don A. Moore.** 2015. "Does the Better-than-Average Effect Show That People Are Overconfident?: Two Experiments." *Journal of the European Economic Association* 13 (2): 293–329.

**Bracha, Anat, and Donald J. Brown.** 2012. "Affective Decision Making: A Theory of Optimism Bias." *Games and Economic Behavior* 75 (1): 67–80.

**Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper.** 2010. "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion." *Econometrica* 78 (4): 1375–1412.

**Brunnermeier, Markus K., and Jonathan A. Parker.** 2005. "Optimal Expectations." *American Economic Review* 95 (4): 1092–1118.

**Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini.** 2013. "Overconfidence and Social Signaling." *Review of Economic Studies* 80 (3): 949–83.

**Camerer, Colin, and Dan Lovallo.** 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review* 89 (1): 306–18.

**Carlson, Ryan W., Michel André Maréchal, Bastiaan Oud, Ernst Fehr, and Molly J. Crockett.** 2020. "Motivated Misremembering of Selfish Decisions." *Nature Communications* 11 (1): 1–11.

**Carrillo, Juan D., and Thomas Mariotti.** 2000. "Strategic Ignorance as a Self-Disciplining Device." *Review of Economic Studies* 67 (3): 529–44.

**Charness, Gary, Aldo Rustichini, and Jeroen Van de Ven.** 2018. "Self-Confidence and Strategic Behavior." *Experimental Economics* 21 (1): 72–98.

**Cheng, Ing-Haw, Sahil Raina, and Wei Xiong.** 2014. "Wall Street and the Housing Bubble." *American Economic Review* 104 (9): 2797–2829.

**Chew, Soo Hong, Wei Huang, and Xiaojian Zhao.** 2020. "Motivated False Memory." *Journal of Political Economy* 128 (10): 3913–39.

**Compte, Olivier, and Andrew Postlewaite.** 2004. "Confidence-Enhanced Performance." *American Economic Review* 94 (5): 1536–57.

**Coutts, Alexander.** 2019. "Good News and Bad News Are Still News: Experimental Evidence On Belief Updating." *Experimental Economics* 22 (2): 369–95.

**de la Rosa, Leonidas Enrique.** 2011. "Overconfidence and Moral Hazard." *Games and Economic Behavior* 73 (2): 429–51.

**DellaVigna, Stefano, and Ulrike Malmendier.** 2006. "Paying Not to Go to the Gym." *American Economic Review* 96 (3): 694–719.

**Di Tella, Rafael, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman.** 2015. "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism." *American Economic Review* 105 (11): 3416–42.

**Eil, David, and Justin M. Rao.** 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself." *American Economic Journal: Microeconomics* 3 (2): 114–38.

**Fang, Hanming, and Giuseppe Moscarini.** 2005. "Morale Hazard." *Journal of Monetary Economics* 52 (4): 749–77.

**Foschi, Matteo, and Luís Santos-Pinto.** 2017. "Subjective Performance Evaluation of Employees with Biased Beliefs." Unpublished.

**Gervais, Simon, and Itay Goldstein.** 2007. "The Positive Effects of Biased Self-Perceptions in Firms." *Review of Finance* 11 (3): 453–96.

**Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia, and Roel van Veldhuizen.** 2015. "Motivated Self-Deception and Unethical Behavior." Unpublished.

**Gottlieb, Daniel.** 2010. "Will You Never Learn? Self-Deception and Biases in Information Processing." Unpublished.

**Gottlieb, Daniel.** 2014 "Imperfect Memory and Choice under Risk." *Games and Economic Behavior* 85: 127–58.

**Haisley, Emily C., and Roberto A. Weber.** 2010. "Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior." *Games and Economic Behavior* 68 (2): 614–25.

**Heidhues, Paul, Botond Kőszegi, and Philipp Strack.** 2018. “Unrealistic Expectations and Misguided Learning.” *Econometrica* 86 (4): 1159–1214.

**Hestermann, Nina, and Yves Le Yaouanq.** 2021. “Experimenting with Self-Serving Attribution Biases.” *American Economic Journal: Microeconomics* 13 (3):198–237.

**Hoffman, Mitchell.** 2016. “How Is Information Valued? Evidence from Framed Field Experiments.” *Economic Journal* 126 (595): 1884–1911.

**Hoffman, Mitchell, and Stephen V. Burks.** 2020. “Worker Overconfidence: Field Evidence and Implications for Employee Turnover and Firm Profits.” *Quantitative Economics* 11 (1): 315–48.

**Huffman, David, Collin Raymond, and Julia Shvets.** 2022. “Replication Data for: Persistent Overconfidence and Biased Memory: Evidence from Managers.” American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E166801V1.

**Hvide, Hans K.** 2002. “Pragmatic Beliefs and Overconfidence.” *Journal of Economic Behavior and Organization* 48 (1): 15–28.

**Kőszegi, Botond.** 2006. “Ego Utility, Overconfidence, and Task Choice.” *Journal of the European Economic Association* 4 (4): 673–707.

**Kruger, Justin, and David Dunning.** 1999. “Unskilled and Unaware of it: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments.” *Journal of Personality and Social Psychology* 77 (6): 1121–34.

**Malmendier, Ulrike, and Geoffrey Tate.** 2005. “CEO Overconfidence and Corporate Investment.” *Journal of Finance* 60 (6): 2661–2700.

**Malmendier, Ulrike, and Geoffrey Tate.** 2015. “Behavioral CEOs: The Role of Managerial Overconfidence.” *Journal of Economic Perspectives* 29 (4): 37–60.

**Merkle, Christoph, and Martin Weber.** 2011. “True Overconfidence: The Inability of Rational Information Processing to Account for Apparent Overconfidence.” *Organizational Behavior and Human Decision Processes* 116 (2): 262–71.

**Mijović-Prelec, Danica, and Drazen Prelec.** 2010. “Self-Deception as Self-Signalling: A Model and Experimental Evidence.” *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365 (1538): 227–40.

**Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat.** 2011. “Managing Self-Confidence: Theory and Experimental Evidence.” NBER Working Paper 17014.

**Niederle, Muriel, and Lise Vesterlund.** 2007. “Do Women Shy Away from Competition? Do Men Compete Too Much?” *Quarterly Journal of Economics* 122 (3): 1067–1101.

**Oster, Emily, Ira Shoulson, and E. Ray Dorsey.** 2013. “Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease.” *American Economic Review* 103 (2): 804–30.

**Park, Young Joon, and Luís Santos-Pinto.** 2010. “Overconfidence in Tournaments: Evidence from the Field.” *Theory and Decision* 69 (1): 143–66.

**Quattrone, George A., and Amos Tversky.** 1984. “Causal versus Diagnostic Contingencies: On Self-Deception and on the Voter’s Illusion.” *Journal of Personality and Social Psychology* 46 (2): 237–48.

**Santos-Pinto, Luís.** 2008. “Positive Self-Image and Incentives in Organisations.” *Economic Journal* 118 (531): 1315–32.

**Santos-Pinto, Luís.** 2010. “Positive Self-Image in Tournaments.” *International Economic Review* 51 (2): 475–96.

**Santos-Pinto, Luís, and Joel Sobel.** 2005. “A Model of Positive Self-Image in Subjective Assessments.” *American Economic Review* 95 (5): 1386–1402.

**Sarver, Todd.** 2018. “Dynamic Mixture-Averse Preferences.” *Econometrica* 86 (4): 1347–82.

**Schwardmann, Peter, and Joël Van der Weele.** 2019. “Deception and Self-Deception.” *Nature Human Behaviour* 3 (10): 1055–61.

**Smith, Adam.** 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Strahan and Cadell.

**Van den Steen, Eric.** 2004. “Rational Overoptimism (and Other Biases).” *American Economic Review* 94 (4): 1141–51.

**Zimmermann, Florian.** 2020. “The Dynamics of Motivated Beliefs.” *American Economic Review* 110 (2): 337–61.