Overconfidence and diversification

Yuval Heller

Nuffield College & Department of Economics, University of Oxford

30. September 2011

Overconfidence and Diversification

Yuval Heller

Nuffield College and the Department of Economics, University of Oxford.
New Road, Oxford OX1 1NF, UK. Phone: +44-1865-278993, +44-7582-540427.

Abstract

Experimental evidence suggests that people tend to be overconfident in the sense that they overestimate the accuracy of their private information, judgment and intuition. In this paper we present a novel evolutionary foundation for overconfidence: diversification of risk. In addition, the model explains various stylized facts that characterize overconfidence. Finally, an equivalent formulation of the model illustrates why principals may prefer overconfident agents in various strategic (non-evolutionary) interactions.

Key words: overconfidence, diversification, evolution.

1 Introduction

In many experimental studies participants are asked to answer trivia questions, and to report the level of confidence (subjective probability) that they answered each of these questions correctly. The typical result in such experiments is that people are overconfident: their confidence systematically exceeds the true accuracy (see Section 2). That is, people tend to overestimate the accuracy of their private information, personal judgment and intuition. Various evidence suggest that overconfidence substantially influences economic behavior of analysts (Friesen and Weller, 2006), investors (e.g., Berber and Odean, 2001), entrepreneurs (e.g., Cooper, Woo and Dunkelberg, 1988; Koellinger, Minniti and Schade,

Email address: yuval.heller@economics.ox.ac.uk or yuval26@gmail.com (Yuval Heller).

[1] This work is in partial fulfillment of the requirements for Ph.D. degree at Tel-Aviv University, and it was supported by the Israel Science Foundation (grant number 212/09). I would like to thank Eilon Solan for his careful supervision, and for the continuous help he has offered. I would also like to express my deep gratitude to Eddie Dekel, Tzachi Gilboa, Zvika Neeman, Ariel Rubinstein, Assaf Romm, Ran Spiegler, Roee Teper, and the seminar participants at Tel-Aviv University, Hebrew University of Jerusalem, Israel Institute of Technology, Haifa University, Ben-Gurion University, Ecole Polytechnique, University of Southampton, Universitat Pompeu Fabra, and University of Oxford, for many useful comments, discussions and ideas. An earlier version of this paper was called "Overconfidence and risk dispersion".

2007), managers (e.g., Rabin and Schrag, 1999; Goel and Thakor, 2008; Gervais, Heaton and Odean, 2010), and consumers (Grubb, 2009).

In this paper, we present a theoretical model that studies the relation between overconfidence and risk diversification. The main application of the model is a novel evolutionary foundation for overconfidence. We show that overconfidence is a unique evolutionary stable behavior, and we characterize its optimal level. Our model has two key properties that distinguish it from existing evolutionary models for overconfidence : (1) the evolutionary dynamics is based only on individual selection (and not on group selection as in Bernardo and Welch, 2001), and (2) overconfidence achieves the optimal solution to the evolutionary problem (and not a "second-best" adaptation that compensates for another bias as in Blume and Easley, 1992; Waldman, 1994; and Wang, 2001).

An equivalent representation of the model describes the strategic interaction between a risk-averse principal and privately informed agents, and illustrates why the principal prefers overconfident agents over rational agents. This may help understanding why it seems that successful analysts, entrepreneurs and managers tend to have high levels of overconfidence.

1.1 Model and Results

The equivalent strategic representation allows us to illustrate our model and to intuitively explain our results in a simpler manner than the evolutionary representation. Therefore, we first describe the strategic representation, and postpone the presentation of the evolutionary framework to the next subsection. Our model describes the strategic interaction between a risk-averse principal and privately informed agents.[2] As an illustrating example, consider the following strategic interaction in a venture capital fund.

Example 1 *A risk-averse CEO of a venture capital fund hires several analysts. When the CEO chooses which analysts to hire, he observes the confidence level of each candidate. Each analyst manages the investments of the fund in his area. For simplicity, assume that each analyst investigates several startup companies, and chooses one of these companies. The analyst may base his choice on two methods: (1) follow the accepted guidelines and make the choice a typical analyst would in such a situation, or (2) be original and follow his own personal judgment and intuition. The fund invests money in the chosen startup company. This investment may either succeed or fail. Successes of different agents are positively correlated if both analysts follow the accepted guidelines, and they are independent otherwise. The analyst is interested in maximizing the success probability of his investment. The CEO wishes to maximize the total number of successful investments, and each additional success has a smaller marginal payoff.*

[2] In the evolutionary framework (as discussed in the next subsection) the agents are the individuals in the population, and the evolutionary dynamics "behaves" as if it were a risk-averse principal.

Our model (Section 3) includes a principal and many agents. Each agent i is characterized by bias function g_i that determines how he evaluates the accuracy of personal judgment and intuition: if his judgment is correct with probability $0 < p_i < 1$, the agent believes the accuracy to be $g_i(p_i)$. The strategic interaction includes two stages. At stage 1 the principal observes the bias functions of potential agents, and chooses which agents to hire. At stage 2, all agents publicly receive signal $0 < q < 1$ - the success probability of following the accepted guidelines.[3] [4] In addition, each agent i privately receives signal $0 < p_i < 1$ (evaluated as $g_i(p_i)$) - the success probability of following his own judgment (each p_i is independently drawn from the same distribution.). Then each agent chooses whether to follow the accepted guidelines or to follow his own judgment. Each agent who follows the accepted guidelines succeeds with probability q; success of different agents who follow the accepted guidelines are positively correlated with correlation coefficient $0 < \rho \leq 1$. Each agent i who follows his own judgment succeeds with probability p_i independent of other agents, An agent receives high payoff if he succeeds and low payoff if he fails. The payoff of the principal is a concave increasing function of the total number of successful agents.[5]

Following the accepted guidelines bears a larger aggregate risk due to the positive correlation between the successes of different agents. This creates a conflict of interests between calibrated agents ($g_i(p_i) = p_i$) who maximize their probability of success, and the risk-averse principal. The principal has a tradeoff between two objectives: (1) maximizing the expected number of successes, and (2) reducing the variance in the number of successes. The first goal is fully consistent with the interests of calibrated agents. However, due to the second goal, the principal would like agents with p_i a little bit smaller than q to follow their somewhat less accurate personal judgment, in order to reduce the variance and achieve a better diversification of risk among the agents.

Our first result (Theorem 3) shows that if the number of agents is sufficiently large, then this conflict is optimally resolved by hiring overconfident agents.[6] That is, there is a continuous and increasing bias function g^*, which always overestimates the perceived accuracy of agent's personal judgment ($g^*(p) > p$ for every $0 < p < 1$, see Figure 1 in Section 3), such that if all agents have this bias function, it approximately induces the "first-best" outcome for the principal - the outcome he would achieve if he could receive all the private signals and directly control the actions of all agents. We further show (Theorem 4) that g^* is unique in the following sense: all other bias profiles, including heterogeneous profiles in which agents have different bias functions, induce strictly worse

[3] We assume that all agent who follow the accepted guidelines have the same success probability q in order to simplify the presentation of the results and make the model more tractable. The results would remain qualitatively similar if each agent who follows the accepted guidelines has an independent success probability q_i.

[4] The assumption that the public signal is evaluated without a bias is without loss of generality as discussed in Subsection 6.2.

[5] In addition, we make the technical assumption that this utility has decreasing absolute risk aversion - see Subsection 3.1.

[6] The conflict can also be resolved by using monetary incentives. In Section 5 we demonstrate why in some situations the "overconfidence" mechanism may be easier to implement than a mechanism that is based on monetary incentives.

outcomes. Our third result (Theorem 6) presents interesting comparative statics. It shows that the principal prefers more overconfident agents if: (1) he becomes more risk-averse, or (2) the correlation coefficient ρ becomes larger. The intuition of both results is that both changes deepen the conflict of interests between the principal and calibrated agents, and more overconfidence is required to compensate it. As demonstrated in Example 5, when the number of agents is small, our results do not hold.

1.2 Evolutionary Application

We now present the main application of our model. Consider a large population of agents with several genetic types. Each type i induces a (possibly random) bias function g_i for its members. In each generation, each agent faces an important decision that influences his fitness. When taking this decision the agent may either follow the accepted guidelines (do what most people think to be best in this situation) or follow his own judgment and take an original action. At the beginning of each generation each agent i receives two signals $0 < p_i < 1$ and $0 < q < 1$ with the same interpretation as in the basic model: p_i is the independent success probability of following agent's judgment, and q is the positively correlated success probability of all agents that follow the accepted guidelines.

In each generation, each agent chooses whether to follow his own judgment or the accepted guidelines, and this leads either to a success or to a failure in terms of fitness (expected number of offspring). The size of each type in the next generation is determined by *replicator dynamics* (the new size of each type in the next generation is their size in the previous generation multiplied by their average fitness) with a small mutation rate (each individual in the new generation has a small probability to be randomly assigned into a new type.).

We are interested in the following question: which type will survive in the long run? A simple adaptation of existing results in the evolutionary literature shows that with high probability in the long run a unique type prevails over the entire population: the type that maximizes the logarithm of the average fitness in each generation.[7] That is, the limiting behavior that is induced by the evolutionary dynamics is the bias profile that is directly chosen by a risk-averse principal with a logarithmic utility function.

The fact that the "principal's" utility is logarithmic,[8] allows us to characterize additional comparative statics on the optimal level of overconfidence. First, we show (Theorem 7) that the optimal level of overconfidence is higher if there is a larger difference between the fitness in case of a success and the fitness in case of a failure. This result is in accordance with the experimental finding of Sieber (1974) which suggests that people tend to be more overconfident with respect to more important decisions (experimental findings and stylized facts are discussed in Subsection 2.2).

[7] In Section 4 we formally adapt Robson (1996)'s result to our setup. See also related results in Lewontin and Cohen (1969), McNamara (1995), and the financial paper of Samuelson (1971).
[8] It is enough to assume that the principal has a constant relative risk aversion utility.

Theorem 7 also shows that the optimal level of overconfidence is higher if the success probabilities tend to be lower (first-order stochastic dominance). This result is in accordance with a stylized fact about overconfidence - the *hard-easy effect* (Lichtenstein, Fischhoff, and Phillips, 1982). According to this effect, the more difficult the task, the greater the observed overconfidence.

Finally, Theorem 7 shows that when the potential gain is large enough and ρ and p_i are close enough to 1, then the perceived error probability of personal judgment $(1 - g^*(p_i))$ is much smaller then the true error probability $(1 - p_i)$. This fits the experimental stylized fact of the *false certainty effect* (Fischhoff, Slovic, and Lichtenstein, 1977): people are often wrong when they are certain about their private information.

Our model implies that all evolutionary histories induce overconfidence, but that there is a large variety in the level of induced overconfidence given different histories. This implication is in accordance with the experimental findings about levels of overconfidence in different cultures, as surveyed in Yates et al. (2002).

1.3 Related Literature

Most related work is the literature studying evolutionary foundations for overconfidence. We discuss this literature in the following paragraphs. In Section 2 we discuss other related papers.

Waldman (1994) showed that "second-best" adaptations can be evolutionarily stable with sexual inheritance, even though they need not be the optimal solution to the evolutionary problem. In particular, He demonstrated that the combination of overconfidence (overestimating self-ability) with excess disutility from effort is a "second-best" adaptation. Similarly, Blume and Easley (1992) and Wang (2001) presented models in which the combination of overconfidence and excess risk aversion (or too high discount factor) are "second-best" adaptations. Contrary to that, in our model overconfidence induces the optimal outcome, and does not compensate for other errors.

In Bernardo and Welch (2001)'s model a small proportion of individuals are overconfident, while the rest of the population are calibrated. Being overconfident reduces the fitness of the individual, but it substantially improves the fitness of his group, by inducing positive information externality in a *cascade* interaction. Under the assumption that the evolutionary dynamics combines both group and individual selection, the evolutionarily stable profile includes a minority of overconfident individuals. Contrary to that, our model only relies on individual selection, and it does does not include information externalities.

Recently, Louge (2010) independently presented a closely-related evolutionary model, and he showed that the evolutionary stable behavior has two overconfidence-related properties: (1) a bias towards actions that defy "common wisdom", and (2) more extreme public information is required before disregarding private information. Louge applied this rule

to *cascade* interactions and demonstrated that herds eventually arise, but the probability of herding on the wrong action is lower than with a rational rule. Our model differs from Louge (2010) in a few aspects: (1) we also deal with strategic (non-evolutionary) interactions of economic interest; (2) our model explains various stylized facts about overconfidence; and (3) we allow for partial correlation between agents who follow the public signal, and we allow the private signals to be costly (see Subsection 6.3).

1.4 Structure

Section 3 presents the model and the results (all proofs are given in the appendix). The main evolutionary application of our model is presented in Section 4. Section 5 presents a few more applications of our model, including: (1) the interaction between investors and entrepreneurs, and (2) an example that shows how overconfidence can increase social welfare. Section 6 presents a few variants and extensions of the basic model, and discusses its key assumptions: (1) we relax the assumption that the number of agents is exogenously given, and we show that the principal always prefers to multiply the number of agents; (2) we allow agents to have bias also with respect to the success probability of following the accepted guidelines, and we show that our results essentially remain the same; (3) we extend our results to a setup where private information is costly, and each agent has to privately invest effort in improving the accuracy of his personal judgment; (4) we show that our results also hold when agents are informed experts who recommend the principal which action to choose; and (5) we show that our results hold also when the agents are more risk averse than the principal; and (6) we show that our model can be reformulated to capture overconfidence as underestimating the variance of a continuous noisy signal (as often modeled in finance papers).

2 Related Literature

The term "overconfidence" has been widely used in psychology since the 1960s, and in the economics and finance literature since the 1990s. Google Scholar reports on 876 papers that include this term in their titles and about 40,000 papers that include it anywhere in the text (September 2010). In this section we briefly discuss a small portion of this literature. We first describe the different definitions of overconfidence and the main experimental and empirical findings about it (Subsection 2.1). Next, we describe a few experimentally observed stylized facts about overconfidence (Subsection 2.2). Finally, we describe related economic and financial models which deal with overconfidence (Subsection 2.3).

The interested reader is referred to the following surveys on overconfidence: the classical survey of Lichtenstein, Fischhoff and Phillips (1982), which summarizes overconfidence literature in the 1960s and 1970s; the survey of Griffin and Brenner (2004) that summarizes the theoretical controversies about overconfidence, and the recent survey of Skala (2008).

2.1 Definitions of Overconfidence and Experimental Literature

The term overconfidence has three different definitions in the literature. The most popular definition (and the one used in this paper) describes overconfidence as a systematic calibration bias, for which the assigned probability that the answers given are correct exceeds the true accuracy (see e.g., Oskamp, 1965; Lichtenstein, Fischhoff and Phillips, 1982; Brenner et al., 1996; Dawes and Mulford, 1996). Overconfidence usually emerges when the uncertainty regarding the true answer is generated by the state of knowledge of the agent - *internal uncertainty* (Howell and Burnett, 1978; Kahneman and Tversky, 1982). For example, the question: *"Is Mont Blanc the tallest mountain in Europe?"*. In such cases, the agent has partial cues and private noisy signals about the correct answer, and using his personal judgment and intuition, he tries to evaluates the accuracy of his preferred answer. As implied by the experimental evidence, in such cases agents tend to overestimate their accuracy. When the source of uncertainty is external to the agent (e.g., tossing a coin or the outcome of a future football game) the tendency for overconfidence is substantially weaker (see, e.g., Budescu, and Du, 2007).

The second definition of overconfidence deals with excessive certainty regarding the accuracy of one's beliefs about an uncertain continuous quantity. Researchers examining this effect typically ask their participants questions with numerical answers (e.g., "How long is the Nile River?"), and then have participants estimate (usually 90%) confidence intervals. Overconfidence is measured by the rate of surprises, i.e., the percentage of true values falling outside the confidence intervals. The typical finding (see Lichtenstein, Fischhoff and Phillips, 1982; Russo and Schoemaker, 1992) is that people tend to present substantial overconfidence: 90% confidence intervals contain on average only 50% of the true values.[9]

The third definition of overconfidence describes the phenomenon in which people believe themselves to be *better than average*. A review of this literature can be found in Alicke and Govorun (2005). A typical finding in this literature is the oft-quoted finding of Svenson (1981) that 77% of Swedish subjects felt they were safer drivers than the median. This bias is closely related to overly positive self-evaluations and to over-optimism about the future. Taylor and Brown (1988) report such phenomena to be positively correlated with different criteria of mental health. Recently, Moore (2007) and Benoit and Dubra (2011) suggest that most of the experimental findings of the better than average phenomenon can also be explained by a fully-rational Bayesian model.

Training improves overconfidence but usually only to a limited extent. Russo and Schoemaker (1992) show that asking people job relevant questions reduced overconfidence from 50% to 30% (for 90% confidence interval). Weather forecasters, who typically have several years of experience in assessing probabilities and receiving an immediate feedback, are

[9] People also present overconfidence for 50% confidence intervals and for free-choice intervals, but this overconfidence is substantially smaller (Soll and Klayman, 2004; Teigen and Jorgensen, 2005).

quite well calibrated (Lichtenstein, Fischhoff and Phillips, 1982; and also expert bridge players - see Keren, 1987). Other experts such as physicians and professional traders, present substantial confidence biases (see, e.g., Koehler, Brenner and Griffin, 2002; Glaser, Langer, and Weber, 2010).

Empirical data suggests that people present overconfidence not only in the lab but also in real-life situations. Russo and Schoemaker (1992) report the following example: "*newly hired geologists were wrong much more than their levels of confidence implied. For instance, they would estimate a 40% chance of finding oil, but when ten such wells were actually drilled, only one or two would produce.*" Berber and Odean (2001) show that men, which are generally believed to be more overconfident than women in areas such as finance, trade more excessively than women, and this excess trade substantially reduces net returns. Henrion and Fischhoff (2002) show that scientists systematically underestimate uncertainty in their own measurements of physical constants. Chuang and Lee (2006) empirically evaluate data on prices of firms in NYSE and AMEX during 1963-2001 and find evidence that investors overestimate accuracy of private information. Similarly, Friesen and Weller (2006) present empirical findings that analysts tend to be overconfident. Recently, Grubb (2009) analyzes consumer tariff choices and usage decisions of cellular services, and show that the consumers seem to be overconfident in their ability to estimate their future demand for cellular services. Finally, Ben-David, Graham and Harvey (2010) demonstrated that financial managers overestimate their ability to predict stock market returns.

2.2 Stylized Facts about Overconfidence

The observed overconfidence in experiments usually satisfies a few recurrent properties (or effects). In this subsection we describe the main observed properties.

One of the main findings in the experimental literature is that the degree of overconfidence depends on the difficulty of the task - the *hard-easy effect*. The more difficult the task, the greater the observed overconfidence (Lichtenstein, Fischhoff, and Phillips, 1982; Moore and Healy, 2008). A few papers suggest that the hard-easy effect, and apparent overconfidence in general may be the result of choosing unrepresentative hard questions in experiments (Gigerenzer, Hoffrage, and Kleinbolting, 1991; Juslin, 1994), or regression toward the mean and boundary effects in the presence of unbiased judgmental random errors (Erev, Wallsten and Budescu, 1994; Soll, 1996; Juslin, Winman and Olsson, 2000). Recent experiments demonstrate that people still present overconfidence (and the hard-easy effect), though to a less extent, when representative questions are used (which are randomly sampled from a natural set) and when unbiased judgmental random errors are taken into account in the analysis (see, e.g., Budescu, Wallsten, and Au, 1997; Klayman et al., 1999; Glaser, Langer and Weber, 2010).

Another finding is the *false certainty effect*: people are often wrong when they are certain in their private information. In the experiment of Fischhoff, Slovic, and Lichtenstein

(1977) participants severely underestimated the probability they erred in seemingly easy questions. Specifically, the error probability of 10% of the questions was estimated by the subjects to be extremely low (less than 1:1,000), while the true error probability in these questions was approximately 10%. The participants had sufficient faith in their confidence judgments to be willing to stake money on their validity.

Griffin and Tversky (1992) suggest that many observed patterns of overconfidence can be explained by the *strength-weight effect*: "people focus on the strength or extremeness of the available evidence with insufficient regard for its weight or credence. This mode of judgment yields overconfidence when strength is high and weight is low, and underconfidence when strength is low and weight is high." (Griffin and Tversky, 1992, p. 411).

Some experiments (e.g., Gigerenzer, Hoffrage and Kleinbolting, 1991; Griffin and Tversky, 1992) compare people's confidence in giving correct answers by two methods: (1) each answer is evaluated separately (*case-by-case evaluations*), and (2) after answering several questions, participants are asked to evaluate the frequency of correct answers (*set-based evaluations*). These papers show that people exhibit less overconfidence (or even underconfidence) when evaluating set-based frequencies.

Finally, Sieber (1974) suggests that when the decision is more important, people tend to be more overconfident. In her experiment, two groups of students were compared. Originally, both groups were told that they were taking their mid-term examination. However, when they began the test, one of the groups ("group A") was told that it is not mid-term, but would be used to coach them to mid-term. The two groups had a similar number of correct answers, but group A presented less overconfidence.

2.3 Financial and Economic Models

In this subsection we briefly survey some related financial and economic models that deal with overconfidence. Some papers study motivational reasons for overconfidence. Bénabou and Tirole (2002) present a multiple-self model, in which a rational agent tries to deceive his future self to be overconfident (overestimate his ability), in order to motivate him to undertake more ambitious goals and persist in the face of adversity. Compte and Postlewaite (2004) present a model in which positive emotions can improve performance, and individuals use biases in information processing that enhance their welfare. Köszegi (2006) and Weinberg (2009) model a decision maker who, in addition to having preferences over material outcomes, also derives "ego" utility from positive self-image. In such a setup, moderate overconfidence raises the expected wealth.

Other papers study the evolutionary process that is generated by wealth that flows between investors in an asset market, and investigate the conditions in which overconfidence can survive or even dominate the market. Blume and Easley (1992) and Wang (2001) present models in which investors have a high level of risk aversion (or high discount factor), and overconfident investors can dominate the market due to trading more aggres-

sively in the right way. Gervais and Odean (2001) show how the tendency for a trader to take too much credit for successes leads relatively-inexperienced successful traders to become overconfident. In markets where inexperienced traders continuously enter and old traders die, there will always be overconfident traders, and these traders will tend to control more wealth than their less confident peers. Rabin and Schrag (1999) show how confirmatory bias (the tendency to interpret ambiguous evidence as confirming the current hypothesis) induces overconfidence.

Van den Steen (2004) models "rational overconfidence". Agents have an unbiased random error when evaluating their success probability for each possible action. When such agents face a choice from a set of alternatives, they are more likely to select actions for which they overestimate the probability of success. Thus they will tend to be overconfident about the likelihood of success of the actions they undertake.

A few papers study the influence of overconfident agents on different markets. Odean (1998) shows that overconfidence among investors in financial markets increases expected trading volume, increases market depth, and decreases the expected utility of overconfident trader. Sandroni and Squintani (2007) show the the presence of some overconfident agents qualitatively change the equilibrium behavior and the policy implications in insurance markets with asymmetric information.

3 Model and Results

3.1 Model

3.1.1 Parameters of the Model

Our model includes seven parameters: $(I, \rho, f_{\mathbf{q}}, f_{\mathbf{p}}, L, H, h)$ where:

- $I = \{1, ..., n\}$ is a set of agents. A typical agent is denoted by i or j.

Each agent faces a choice between two actions: (1) $a_{guidelines}$, which is interpreted as following the accepted guidelines, or doing what a typical agent would do in such a situation; and (2) $a_{original}$, which is interpreted as following the personal judgment and intuition of the agent, and making an original choice. Each agent may either succeed or fail, depending on his chosen action and on the state of nature.

- The number $0 < \rho < 1$ is the correlation coefficient between the success of two agents who follow the accepted guidelines. If at least one of the agents followed his own judgment, then their successes are independent and uncorrelated.
- Distribution $f_{\mathbf{q}}$ is a continuous pdf (probability density function) with full support: $\forall 0 < q < 1,\ f_{\mathbf{q}}(q) > 0$.[10] The success probability of all agents who follow the accepted

[10] The full support assumption is given to simplify the presentation of the results. The results

guidelines is the random variable $\mathbf{q}$, which is distributed according to $f_{\mathbf{q}}$.

- Distribution $f_{\mathbf{p}}$ is a continuous pdf with full support. The success probability of each agent i who follows his own judgment is the random variable $\mathbf{p}_i$, which is distributed according to $f_{\mathbf{p}}$.
- $L, H \in \mathbb{R}$ $(H > L)$ are the payoffs an agent may obtain: success yields high payoff (H) and failure yields low payoff (L). In the strategic (non-evolutionary) interactions (such as, Example 1) one can assume that $H = 1$ and $L = 0$. In the evolutionary application described in the next section, H (L) is the fitness of a successful (unsuccessful) agent.
- $h : [L, H] \to \mathbb{R}$ is a strictly concave increasing function that satisfies decreasing absolute risk aversion (DARA). That is: (1) $h' > 0$, (2) $h'' < 0$, and (3) Arrow-Pratt coefficient of absolute risk aversion $r_A(x) = -\frac{h''(x)}{h'(x)}$ is decreasing in x.[11] The function h is interpreted as the utility of the risk-averse principal (described below).

3.1.2 State of Nature

The unknown state of nature determines the value of the tuple of random variables

$$\left(\mathbf{q}, (\mathbf{p}_i)_{i\in I}, \xi_q, (\xi_{i,p}, \xi_{i,q})_{i\in I}\right) \in \left([0,1] \times [0,1]^I \times \{0,1\} \times (\{0,1\}, \{0,1\})^I\right) :$$

- As described earlier, $\mathbf{q} \sim f_{\mathbf{q}}$ and each $\mathbf{p}_i \sim f_{\mathbf{p}}$. The variables $\left(\mathbf{q}, (\mathbf{p}_i)_{i\in N}\right)$ are independent.
- Variable ξ_q is equal to 1 with probability $\mathbf{q}$ (and 0 otherwise). When$\xi_q = 1$ $(\xi_q = 0)$ the accepted guidelines are relevant and updated (irreverent or obsolete).
- If $\xi_q = 1$ $(\xi_q = 0$) then each $\xi_{i,q}$ is equal to 1 with high probability: $\sqrt{\rho} + \left(1 - \sqrt{\rho}\right) \cdot \mathbf{q}$ (with low probability: $\left(1 - \sqrt{\rho}\right) \cdot \mathbf{q}$) and equal to 0 otherwise. Following the accepted guidelines would yield agent i high (low) payoff when$\xi_{i,q} = 1$ $(\xi_{i,q} = 0)$.

Observe that without conditioning on the value of ξ_q the success probability of following the accepted guidelines is $\mathbf{q}$. That is, for each $q \in [0,1]$: m

$$P(\xi_{i,q} = 1 | \mathbf{q} = q) = (1-q) \cdot (1 - \sqrt{\rho}) \cdot q + q \cdot (\sqrt{\rho} + (1 - \sqrt{\rho}) \cdot q) = q.$$

Also observe that conditionally on $\mathbf{q}$, the correlation coefficient between the successes of each two agents i, j who follow the public signal is ρ. That is, for each $q \in [0,1]$:

---

are qualitatively unaffected by relaxing this assumption.

[11] The DARA assumption is not required for Theorem 3 (parts 1,2 and 4) and for Theorem 4. It is required for part 3 of Theorem 3 and for the comparative statics (Theorem 6).

$$\begin{aligned}\rho\left(\xi_{i,q},\xi_{j,q}|\mathbf{q}=q\right) &= \frac{\mathbf{E}\left(\xi_{i,q}\cdot\xi_{j,q}|\mathbf{q}=q\right)-\mathbf{E}\left(\xi_{i,q}|\mathbf{q}=q\right)\cdot\mathbf{E}\left(\xi_{j,q}|\mathbf{q}=q\right)}{\sqrt{var\left(\xi_{q,i}|\mathbf{q}=q\right)\cdot var\left(\xi_{q,i}|\mathbf{q}=q\right)}} \\ &= \frac{q\left(\sqrt{\rho}+\left(1-\sqrt{\rho}\right)\cdot q\right)^2+(1-q)\left(\left(1-\sqrt{\rho}\right)\cdot q\right)^2-q^2}{\sqrt{q\left(1-q\right)q\left(1-q\right)}} \\ &= \frac{q\rho+\left(\left(1-\sqrt{\rho}\right)\cdot q\right)^2+2q^2\sqrt{\rho}\left(1-\sqrt{\rho}\right)-q^2}{q\left(1-q\right)} \\ &= = \frac{\left(1-\sqrt{\rho}\right)q^2\left(1+\sqrt{\rho}\right)+q\rho-q^2}{q\left(1-q\right)} = \frac{-\rho q^2+q\rho}{q\left(1-q\right)} = \rho.\end{aligned}$$

- For each $i\in I$, $\xi_{i,p}$ is equal to 1 with probability $\mathbf{p}_i$ (and 0 otherwise). When $\xi_{i,p}=1$ ($\xi_{i,p}=0$) the personal judgment of agent i is correct (incorrect), and following it would yield agent i high (low) payoff.

Variables $(\xi_{i,q},\xi_{i,p})_{i\in N}$ are independent conditionally on $(\mathbf{q},\xi_q)$, and variables $\left(\xi_q,(\xi_{i,p})_{i\in N}\right)$ are independent.

3.1.3 Strategic Interaction

The strategic interaction between the principal and the agents includes two stages. At stage 1 the principal (who has no information on the state of nature) chooses a profile of bias functions $(g_i)_{i\in I}$. Each function $g_i:[0,1]\to[0,1]$ determines the bias of agent i when estimating the accuracy of his own judgment. That is, if the true success probability of following personal judgment is p_i, then agent i mistakenly believes it to be $g_i\left(p_i\right)$.[12] The choice of the bias profile $(g_i)_{i\in I}$ is interpreted as follows: there is an infinite pool of potential agents with all possible bias functions. The principal can observe these biases, and choose $|I|$ agents with any given profile of bias functions.[13] The principal is fully rational and knows all aspects of the model. Each agent has bounded rationality, and he is not aware that he has a confidence bias.[14]

Remark 2 The assumption that the principal can choose the bias profile fits well the evolutionary application (described in the next section). The assumption is very simplistic when considering strategic interactions (such as, Example 1). However, the intuition and main implications of our model can also be applied in a more complicated setup where the principal cannot choose the optimal bias profile, but can only approximate it by choosing a bias profile from a finite set of feasible profiles (e.g., the principal meets several potential agents, and observes a noisy signal about their confidence biases).

[12] See Subsection 6.2 for discussing and relaxing the assumption that agents are biased only with respect to personal judgment.

[13] The number of agents the principal hires is exogenously given in the basic model. In Subsection 6.1 we extend the model to allow the principal to choose the number of hired agents.

[14] Such an assumption is in accordance with the findings of Friesen and Weller (2006) which suggest that: (1) analysts are overconfident; (2) an analyst is not aware of his own bias; and (3) an analyst is aware that other analysts tend to be overconfident.

After stage 1, all agents are publicly informed about the value of $\mathbf{q}$ (the success probability of following the accepted guidelines), and each agent i with bias function g_i, is privately misinformed that the value of $\mathbf{p}_i$ is $g_i(\mathbf{p}_i)$.

At stage 2 each agent i chooses an action $a_i \in \{a_{guidelines}, a_{original}\}$, where $a_{guidelines}$ ($a_{oriignal}$) is interpreted as following the accepted guidelines (personal judgment). The payoff of agent i is as follows:

$$\mathbf{u}_i(a_{guidelines}) = \begin{cases} H & \text{if } \xi_{i,q} = 1 \\ L & \text{if } \xi_{i,q} = 0 \end{cases}, \text{ and } \mathbf{u}_i(a_{original}) = \begin{cases} H & \text{if } \xi_{i,p} = 1 \\ L & \text{if } \xi_{i,p} = 0 \end{cases},$$

Our assumption that f_p and f_q are continuous guarantee that the inequality $\mathbf{q} \neq g(\mathbf{p}_i)$ holds with probability 1. Thus, each bias profile $(g_i)_{i\in I}$ induces a strictly dominating strategy profile for each agent i: following the accepted guidelines if $\mathbf{q} > g_i(\mathbf{p}_i)$, and following the personal judgment if $\mathbf{q} < g_i(\mathbf{p}_i)$.[15] Let $\mathbf{u}_i(g_i) = u_i(g_i, \mathbf{p}_i, \mathbf{q}, \xi_q, \xi_{i,p}, \xi_{i,q})$ be the random payoff of agent i who uses this strictly dominating strategy.

The payoff of the principal, $u\left((g_i)_{i\in I}\right)$, is a strictly concave increasing vN-M (von-Neumann and Morgenstern, 1944) function of the average payoff of the agents (or, equivalently, of the number of successful agents):

$$u\left((g_i)_{i\in I}\right) = \mathbf{E}_{(\mathbf{p}_i)_{i\in I}, \mathbf{q}, \left(\xi_{i,\mathbf{p}}, \xi_{\mathbf{i},\mathbf{q}}\right)_{i\in I}} \left(h\left(\frac{1}{n}\sum_{i\in I} \mathbf{u}_i(g_i)\right)\right).$$

That is, the principal wishes to maximize the total number of successes, and each additional success has a smaller marginal payoff.

3.1.4 Auxiliary Definitions

Bias profile $(g_i^*)_{i\in I}$ is ϵ-optimal (for $\epsilon > 0$) if it yields the best payoff up to ϵ: $u\left((g_i^*)_{i\in I}\right) > u\left((g_i)_{i\in I}\right) - \epsilon$ for every profile $(g_i)_{i\in I}$. Let the first-best payoff of the game, be the payoff that can be achieved by the principal when he obtains all the private signals $(\mathbf{p}_i)$ and has full control over the actions of the agents. A bias profile ϵ-induces the first-best payoff if its payoff is as good as the first-best payoff up to ϵ.

Bias profile $(g_i)_{i\in I}$ is homogeneous (or symmetric) if all agents have the same bias function: $\forall i, j \in I,\ g_i = g_j$. With some abuse of notations, we identify a function $g : [0,1] \to [0,1]$ with the homogeneous profile $(g)_{i\in I}$. We say that g is an *optimal bias function* (for a large number of agents) if, for every $\epsilon > 0$, there is large enough n_0 such that, for any game with at least n_0 agents, g is an ϵ-optimal profile. Similarly, we say that g *induces the first-best payoff* (for a large number of agents) if for every $\epsilon > 0$, there is large enough n_0 such that, for any game with at least n_0 agents, g ϵ-induces the first-best payoff.

[15] Playing arbitrary if $\mathbf{q} = g(\mathbf{p}_i)$ (a 0-probability event).

Bias profile $(g_i)_{i\in I}$ is *heterogeneous* if there is a set $Q \subseteq [0,1]$ with a positive Lebesgue measure such that, for each $q \in Q$, $\min_i (g_i)^{-1}(q) < \max_i (g_i)^{-1}(q)$. With some abuse of notation, we identify the bias profile $(g_i)_{i\in I}$ with the following bias profile in a game with $k \cdot |I|$ agents: agents $\{1, ..., k\}$ have bias function g_1, agents $\{k+1, ..., 2k\}$ have bias function g_2, ..., and agents $\{k \cdot (|I|-1)+1, ..., k|I|\}$ have bias function $g_{|I|}$.

3.2 Main Results

The following theorem characterizes the optimal bias function (all proofs are given in the appendix). It shows that there exists a unique optimal bias function g^* that reveals overconfidence: $g^*(p) > p$ for every $0 < p < 1$. Moreover, this overconfidence bias induces the principal's first-best payoff.

Theorem 3 *There exists a unique optimal bias function g^*, which induces the first-best payoff, with the following properties:*

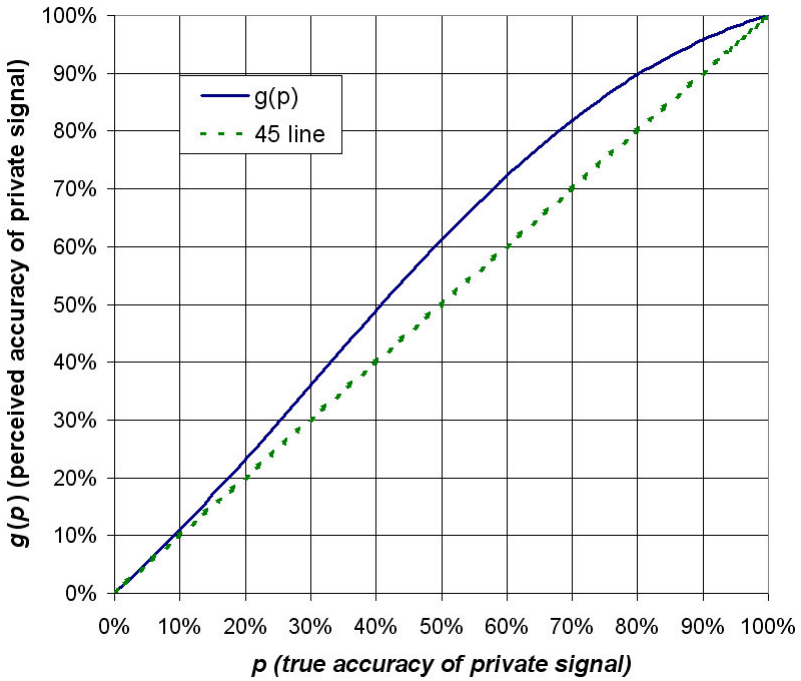
(1) Overconfidence: $g^*(p) > p$ for every $0 < p < 1$.
   (a) g^* is continuous.
   (b) g^* is increasing: $\frac{dg^*(p)}{dp} > 0$ for every $0 < p < 1$, $g^*(0) = 0$, and $g^*(1) = 1$.
   (c) g^* does not depend on the distribution f_q.

The intuition for Theorem 3 is as follows. There is a conflict of interest between calibrated agents ($g_i(p_i) = p_i$) who maximize their probability of success, and the principal who wishes some agents with $\mathbf{p}_i < \mathbf{q}$ to follow personal judgment in order to achieve better risk diversification and to reduce the variance of the fraction of successes. The optimal action of agent i in the principal's first-best profile generally depends on the entire realized profile of signals: $(\mathbf{p}_1, .., \mathbf{p}_n, \mathbf{q})$. However, when there are many agents, the realized empirical distribution of $(\mathbf{p}_1, .., \mathbf{p}_n)$ is very close to its prior distribution $f_{\mathbf{p}}$. Thus, approximately, the first-best choice of agent i only depends on the realizations of $\mathbf{p}_i$ and $\mathbf{q}$. Specifically, for every q, there is some threshold level $g^{-1}(q) < q$ such that it is approximately optimal for the principal if each agent i follows his personal judgment if and only if $\mathbf{p}_i > g^{-1}(q)$. These thresholds construct the optimal bias function $g(p)$. This optimal level of overconfidence aligns the preferences of the principal and the agents. That is, the agents behave as if they have the payoff function of the principal.

Part 4 of Theorem 3 holds due to our simplifying assumption that all agents have the same success probability when following the accepted guidelines. If agents were facing different values of q when following the accepted guidelines, then the optimal level of overconfidence would also depend on $f_{\mathbf{q}}$.

Figure 1 demonstrates what a typical optimal bias function g^* looks like. The values of the parameters are as follows: $\rho = 1$ (perfect correlation between different agents who follow the accepted guidelines), uniform distribution for the accuracy of the private signal, payoffs are $H = 3$ and $L = 1$, and the principal's utility is logarithmic ($h(x) = ln(x)$).

Figure 1. An Example for an Optimal Confidence-Bias Function
Parameters: $\rho = 1$, $H = 3$, $L = 1$, $h(x) = ln(x)$, $f_p \sim uniform((0,1))$



Theorem 3 shows uniqueness in the set of homogeneous bias profiles. That is, it shows that any other homogeneous bias profile induces a worse outcome than g^*, given that the number of agents is sufficiently large. Theorem 4 extends the uniqueness also to the set of heterogeneous profiles. It shows that every heterogeneous profile can be replaced with an homogeneous profile that induces a strictly better outcome, given that the number of agents is sufficiently large.

Theorem 4 *Let $(g_i)_{i\in I}$ be an heterogeneous profile. Then there is $k_0 \in \mathbb{N}$ such that there is an homogeneous profile that induces a strictly better payoff than $(g_i)_{i\in I}$ in the game with $k \cdot |I|$ agents for every $k \geq k_0$.*

The intuition for Theorem 4 is as follows. Let g be a bias function (an homogeneous profile) that induces the same expected number of agents who follow the public signal as the profile $(g_i)_{i\in I}$ (for every $0 < q < 1$). One can show that on average g induces a strictly higher success probability for agents who follow their personal judgment. If the number of agents is sufficiently large, then the law of large numbers implies that g induces a strictly better payoff.

Example 5 shows that Theorems 3-4 are not valid when the number of agents is small. It demonstrates: (1) an asymmetric bias profile that induces higher payoff than the best bias function; and (2) a first-best outcome which is strictly better than what can be induced by bias profiles.

Example 5 *There are two agents. The low payoff is zero ($L = 0$), the high payoff is one ($H = 1$). There is perfect correlation between agents who follow the accepted guidelines ($\rho = 1$). The distribution of each p_i is uniform in $(0, 0.5)$. The principal's utility $h(x)$ is*

$2x$ *if* $x < 0.5$ *and* 1 *if* $x \geq 0.5$.[16] *That is, the principal wishes that at least one agent succeed, and he does not care whether the other agent also succeeds. Consider the case in which* $\mathbf{q} = 0.7$. *One can see that the best bias function is one such that (approximately)* $g^*(0.34) = 0.7$,[17] *and that it induces payoff 0.75. The principal can achieve a higher payoff of 0.775 by using the following optimal heterogeneous bias profile: one agent always follows the accepted guidelines while the other agent always follows his personal judgment. The principal's first best payoff is even higher - 0.8, and it is achieved by observing both* p_1 *and* p_2, *and choosing that the agent with the higher (lower)* p_i *follows his personal judgment (the accepted guidelines).*

3.3 Characterization and Comparative Statics

Our third result (Theorem 6) presents interesting comparative statics. It shows that the principal hires more overconfident agents if: (1) he becomes more risk-averse, or (2) the correlation coefficient ρ becomes larger.

Theorem 6 *Let* $(I_1, \rho_1, f_{\mathbf{q},1}, f_{\mathbf{p},\mathbf{1}}, L_1, H_1, h_1)$ *and* $(I_2, \rho_2, f_{\mathbf{q},2}, f_{\mathbf{p},2}, L_2, H_2, h_2)$ *be two sets of parameters of our model, and let* g_1^* *(*g_2^**) be the unique optimal bias function given the first (second) set of parameters. Then* g_1^* *presents more overconfidence (*$g_1^*(p) > g_2^*(p)$ *for every* $0 < p < 1$*) in each of the following cases:*

(1) Utility h_1 is more risk-averse than h_2 and all other parameters are the same. That is, $h_1 = \psi \circ h_2$ where ψ is concave and increasing.
(2) The first correlation coefficient is larger and all other parameters are the same: $\rho_1 > \rho_2$.

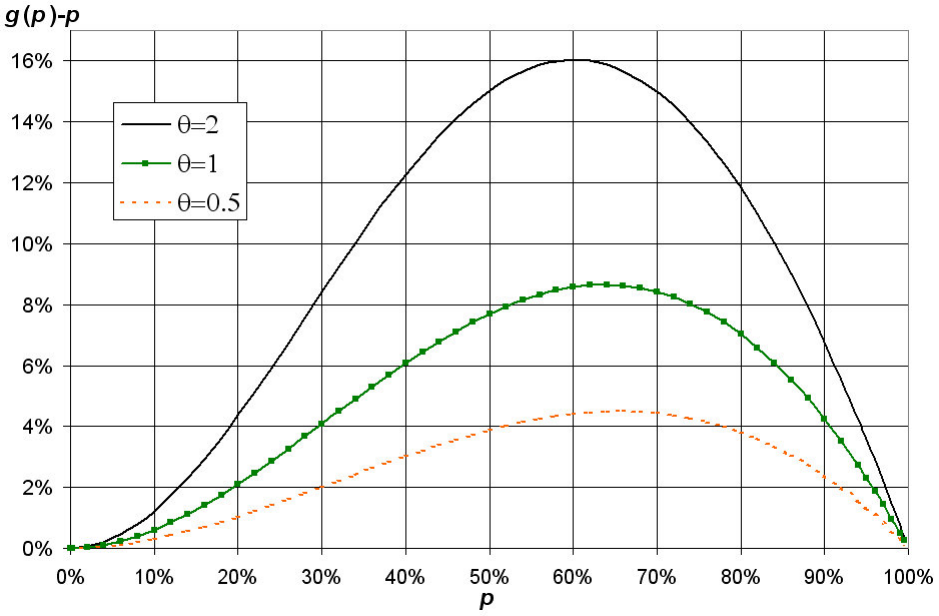
The intuition of Theorem 6 is as follows. If the principal becomes more risk-averse, then he gives more importance to reducing the variance of the number of successes. This deepens the conflict of interest with calibrated agents, and more induced overconfidence is required to align the preferences of the agents and the principal. Similarly, if the correlation coefficient becomes larger, this enlarges the aggregate risk that is induced by following the accepted guidelines, and it deepens the conflict of interest between the principal and calibrated agents.

Figure 2 demonstrates part 1 of Theorem 6. It assumes that the principal's utility has constant relative risk aversion (CRRA, see next subsection) with parameter θ, and it shows the optimal overconfidence bias $(g^*(p) - p)$ for different levels of relative risk aversion: $\theta = 2$, $\theta = 1$ (i.e., $h(x) = ln(x)$), and $\theta = 0.5$. The values of the other parameters in the figure are: $H = 2$, and $L = 1$, and $f_{\mathbf{p}}$ is uniform).

[16] To simplify the example, we use a weakly concave and increasing function h and a distribution $f_{\mathbf{p}}$ without full support. The example can be adapted such that h would be strictly concave and increasing and $f_{\mathbf{p}}$ would have full support.

[17] $(g^*)^{-1}(0.7) = 0.34$ maximizes the expression: $F^2(p_0) \cdot 0.7 + 2 \cdot (1 - F(p_0)) \cdot F(p_0) \cdot (0.7 + 0.3 \cdot \mathbf{E}(p|p > p_0)) + (1 - F(p_0))^2 \left(1 - (1 - \mathbf{E}(p|p > p_0))^2\right)$.

Figure 2. Overconfidence $(g^*(p) - p)$ for different risk aversion levels $(H = 2,\ L = 1,\ f_{\mathbf{p}} \sim \text{Uni}(0,1))$



4 Evolutionary Application

In this section we present the main application of our model, and explain why overconfidence is a unique evolutionary stable behavior.

4.1 Model

Consider a large population of agents with several genetic types: $(T_1, ..., T_K)$. Each type k induces a (possibly random) bias function g_k for its members. In each generation, each agent faces an important decision that influences his fitness. For example, choice of occupation or living area, how to provide food for the family, or how to raise and educate his children. When making a decision the agent may either follow accepted guidelines (do what most people think would be best in his situation) or follow his own judgment and take an original action. At the beginning of each generation each agent i receives two signals $0 < p_i < 1$ and $0 < q < 1$. These signals have the same interpretation as in the basic model: p_i is the independent success probability of following personal judgment, and q is the positively correlated success probability of agents who follow the accepted guidelines.

In each generation, each agent chooses whether to follow his own judgment or follow the accepted guidelines, and this leads either to success or to failure in terms of fitness (number of offspring): success yields high fitness - H and failure yields low fitness - L. The size of each type (the number of its members) in the next generation is determined by *replicator dynamics* with a small positive mutation rate. That is, basically (without regarding the mutation rate) the new size of each type in the next generation is their size in the previous generation multiplied by their average fitness, and their new proportion

in the population is their new size divided by the new total population size. In addition, each individual in the next generation has a small chance to be randomly assigned into a new type.

A well known argument in evolutionary literature (see, Lewontin and Cohen, 1969; McNamara, 1995; Robson, 1996; and the finance-related paper of Samuelson, 1971) shows that with high probability in the long run a unique type prevails over the entire population: the type that maximizes the expectation of the logarithm of the average fitness in each generation. [18]

Formally (adapting the notations of Robson, 1996 to our setup), let $\xi_t = 1$ ($\xi_t = 0$) be the event that the accepted guidelines are correct (incorrect) in generation t (denoted by ξ_q in the previous section). Recall that $P(\xi_t = 1) = q$, and assume that ξ_t-s in different generations are independent. Let $\mathbf{u}_{i,k,t}$ be the fitness of agent i of type k in generation t, and let $\mathbf{m}_k(\xi_t) = \mathbf{E}(\mathbf{u}_{i,k,t}|\xi_t)$ be the expected number of offspring produced by an agent of type k conditional on ξ_t. Robson (1996, Theorem 2-iii) shows that if mutation rates are small, a long time elapsed, and the population is found to have avoided extinction, then the entire population is prevailed by the type that maximizes:

$$\mathbf{E}(ln(\mathbf{m}_k)) = q \cdot ln(m_k(\xi_t = 1)) + (1-q) \cdot ln(m_k(\xi_t = 0)).$$

Observe that due to the law of large numbers, if the population of type k is large enough, then conditional on the value of ξ_t, the realized average fitness of the members of type k in generation t is very close to $\mathbf{m}_k(\xi_t)$: $\mathbf{m}_k(\xi_t) \approx \frac{1}{|T_k|}\sum_{i \in T_k} \mathbf{u}_{i,k,t}$. Thus, Robson (1996)'s result implies that in the long run nature selects the type that maximizes the expectation of the logarithm of the average fitness in each generation. Because of this, the long run limiting behavior that is the result of the evolutionary dynamics can be described as the bias profile that is directly chosen by a risk-averse principal with a logarithmic utility function. [19] Thus, in the long run the homogeneous bias profile g^* is a unique evolutionary stable behavior, and all of the results of Section 3 hold in this setup as well.

In what follows we sketch out the intuition behind this result. Let n_k be the initial number of members of type k, and let $\mathbf{X}_{t,k} = \frac{1}{|T_k|}\sum_{i \in T_k} \mathbf{u}_{i,k,t}$ be the average fitness of type k in generation t. The size of each type T_k after M generations is $n_k \cdot \mathbf{X}_{1,k} \cdot ... \cdot \mathbf{X}_{M,k}$, which is equal to:

$$n_k \cdot \mathbf{X}_{1,k} \cdot ... \cdot \mathbf{X}_{M,k} = n_k \cdot e^{log\left(\mathbf{X}_{1,k} \cdot ... \cdot \mathbf{X}_{M,k}\right)} = n_k \cdot e^{log\left(\mathbf{X}_{1,k}\right) + ... + log\left(\mathbf{X}_{M,k}\right)}.$$

Assume that n_k is large enough and that $\mathbf{E}(log(\mathbf{X}_{t,k})) > 0$, then each $X_{t,k}$ is approxi-

[18] See also Curry (2001) who show that this is equivalent to maximizing the expected relative number of offspring.

[19] See Rayo and Becker (2007) for a discussion why in cases where local and global maxima coincide (as in our setup), one can replace the study of the evolutionary trial-and-error dynamics with an optimization problem of a principal with an appropriate utility function.

mately identically distributed. Assuming that M is large enough then the size of type T_k after M generations is approximately (using the law of large numbers): $n_k \cdot e^{M \cdot \mathbf{E}\left(log\left(\mathbf{X}_{t,k}\right)\right)}$. This depends only on the expectation of the logarithm of the average fitness in each generation, and the type that maximizes this expression will expand exponentially faster than any other type.

At first glance, it might be puzzling that our dynamics is entirely based on individual selection, and yet natural selection does not choose agents who maximize their expected number of children. This is because natural selection "cares" for the number of offspring in the long run. This is not the same as maximizing the "short-run" expected number of children. A calibrated agent has an higher expected number of children than an agent with bias g^* but he also has a higher variance. Generations in which the realized average number of children of calibrated agents is small, substantially reduce the number of offspring in the long run. Due to this, calibrated agents have less offspring in the long run.

4.2 Characterization and Comparative Statics

In the evolutionary application of our model the utility of the "principal" is determined endogenously to be a logarithmic utility function. The specific characteristics of this utility, or more generally of the family of constant relative risk aversion (CRRA) utilities, allows us to further characterize the optimal level of overconfidence. Throughout this subsection we assume that the utility of the principal satisfies CRRA. That is

$$h\left(x\right) = \begin{cases} \frac{x^{1-\phi}}{1-\phi} & \text{if } \phi > 0, \phi \neq 1, \\ ln\left(x\right) & \text{if } \phi = 1, \end{cases}$$

where the parameter $\phi > 0$ specifies the level of (relative) risk aversion.

Let $D = \frac{H-L}{L}$ be the (normalized) potential gain: the ratio between the extra payoff that can be gained when succeeding $(H - L)$ and the minimal guaranteed payoff (L). Theorem 7 shows that CRRA utility yields the following:

(1) The optimal level of overconfidence depends on the payoffs L and H only through its dependency on the potential gain.
(2) Larger potential gain induces more overconfidence. This fits the experimental finding of Sieber (1974), which was discussed in Subsection 2.2.
(3) If the success probability of following personal judgment become smaller (first-order stochastic dominance), then it induces more overconfidence. This fits the experimentally observed *hard-easy effect* (Lichtenstein, Fischhoff, and Phillips, 1982): the more difficult the task, the greater the observed overconfidence (as discussed in Subsection 2.2).
(4) When D is large enough and ρ and p are close enough to 1, then the perceived error probability of personal judgment $(1 - g^*\left(p\right))$ is much smaller then the true error probability $(1 - p)$. This fits the experimentally observed *false certainty effect*

(Fischhoff, Slovic, and Lichtenstein, 1977): people are often wrong when they are certain in their private information (as discussed in Subsection 2.2).

Theorem 7 *Assume that the principal has a CRRA utility function. Then:*

(1) g^* depends on the payoffs only through its dependency on the potential gain D.
(2) If $D_1 > D_2$ (and all other parameters are the same) then g_1^* presents more overconfidence ($g_1^*(p) > g_2^*(p)$ for every $0 < p < 1$).
(3) If distribution $f_{\mathbf{p},2}$ has first order stochastic dominance over $f_{\mathbf{p},\mathbf{1}}$ (and all other parameters are the same) then g_1^* presents more overconfidence.

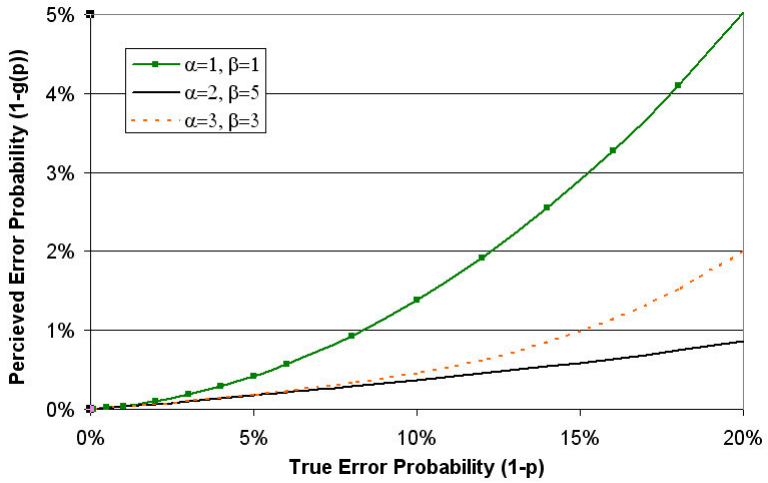
The intuition of the first result is that evaluations of alternatives by a principal with CRRA utility are unaffected by scale, and due to this the optimal bias profile depends only on the normalized potential gain. The intuition of the second result is that larger potential gain, enlarges the aggregate risk of following the accepted guidelines. This deepens the conflict of interest between the principal and calibrated agents, and more overconfidence is required to compensate for it. The intuition of the third result is that the principal wishes that agents with the highest success probabilities would follow their personal judgments. When there is higher probability of receiving lower success probabilities, each accuracy level p_i is more likely to be one of the higher levels.

(1) The ratio between the perceived error probability and the true error probability of personal judgment ($\frac{1-g(p)}{1-p}$) converges to $(1/(D+1))^{\phi} = \left(\frac{L}{H}\right)^{\phi}$ when both p and ρ converge to 1.

The last result (*false certainty effect*) is illustrated in Figure 3. The figure shows the perceived error probability and the true error probability of personal judgment for perfect correlation ($\rho = 1$) and large potential gain $D = 30$, and for three prior beta distributions for the accuracy of the private signals: (1) uniform distribution ($\alpha = 1$, $\beta = 1$, expectation - 50%), (2) single-peaked distribution around 20% ($\alpha = 2$, $\beta = 5$, expectation - 30%), and (3) single-peaked symmetric distribution ($\alpha = 3$, $\beta = 3$, expectation - 50%). The figure demonstrates the false certainty effect, especially for the two single-peaked distributions: when the true error probability is 10% the perceived probability is less than 0.5% (1% for the uniform distribution).

The assumptions that the potential gain is high and the correlation coefficient is close to 1 may seem too extreme. However, one can extend our results into a setup where potential gain $\mathbf{D}$ and correlation coefficient ρ are random variables, and that their joint distribution has some positive small weight on high values. Each type is assumed to induce a single bias function $g^*(p)$ for all values of $\mathbf{D}$ and ρ because either: (1) it is too complicated to induce numerous bias functions $g^*(p|\mathbf{D},\rho)$, or (2) individuals do not know the realization of the potential gain and correlation coefficient when they choose their actions. Observe that for relatively low levels of potential gain $\mathbf{D}$, values of ρ substantially smaller than 1 and low error probabilities (high p-s), the difference in the type's payoff from either choosing personal judgment or accepted guidelines is small (both yield a high payoff). However, when the potential gain $\mathbf{D}$ is high and the correlation coefficient is near 1, the

Figure 3. Perceived vs. True Error Probability ($\rho = 1$, $D = 30$)



chosen action has greater influence on the type's payoff. Thus, for low error probabilities, the single optimal confidence bias function $g^*(p)$ would be close to the value of $g^*(p|\mathbf{D},\rho)$ of high realizations of $\mathbf{D}$ and ρ.

Yates et al. (2002) summarize results from several studies and report that different cultures (in particular, Asian and Western) present overconfidence, but there is a substantial difference in the average level of observed overconfidence. This result can be explained by our model, which predicts that all evolutionary histories would induce overconfidence, but that the optimal level of overconfidence will substantially differ among different societies with different evolutionary histories. In particular, the optimal level of overconfidence depends on: (1) the typical success probability of following personal judgment, (2) the correlation between two agents who follow the accepted guidelines, and (3) the typical potential gain.

5 Applications and Examples

5.1 Strategic Interactions

In this subsection we elaborate and discuss a few examples for the applicability of our model in strategic interactions of economic interest.

5.1.1 CEO and Analysts (Example 1 Revisited)

Recall that Example 1 demonstrates why a risk-averse CEO prefers to hire overconfident analysts, who induce better risk diversification. The CEO could also solve the conflict of interests with calibrated agents by using monetary incentives. However, implementation of such incentives would require the agreement of the firm's shareholders. If the shareholders are risk-natural (for example, due to having a diversified portfolio), they would not approve

such a policy, as they only care for the expected number of successes. On the other hand, the selection of overconfident analysts can be done by the CEO without formally informing the shareholders. It is interesting to compare this result with the model of Gervais, Heaton and Odean (2010), in which, given that the CEO is risk-averse, it is optimal for the risk-neutral shareholders, if the CEO is overconfident, and overestimates his ability to reduce risks. [20]

With minor changes, Example 1 can also describe the following related strategic interactions: (1) each agent is a local distributor and the principal is a global manufacturer, (2) each agent is an editor (or a producer) of a publishing company (or a film studio), and (3) each agent is a researcher in a research and development department of a firm or a non-profit organization. Observe that in the first case (manufacturer and local distributors), competition with other manufacturers may restrict the plausible contracts between the manufacturer and the distributors, and limit the ability of the principal to use monetary incentives to align preferences with the agents.

5.1.2 Investor and Entrepreneurs

Our model can also describe an interesting aspect of the interaction between a risk-averse angel investor (principal) and several entrepreneurs (agents). The entrepreneurs are founders of startup companies who work in a similar area. When the investor interviews the entrepreneurs (before choosing them), he obtains a signal on their confidence-bias. During the development process, each such entrepreneur may either choose a "common" design or an "original" design for his developed product. The successes of different entrepreneurs who chose "common" designs for their products are positively correlated. The conflict of interests between the risk-averse investor and the entrepreneurs can be resolved either by choosing overconfident entrepreneurs or by using monetary incentives. However, the latter method may be too expensive: if each entrepreneur holds a large share of his company, then large monetary incentives are necessary to encourage the choice of "original" designs with smaller success probability.

Our model presents a new explanation why entrepreneurs tend to have high levels of overconfidence (see experimental evidence for this in: Cooper, Woo, and Dunkelberg, 1988; and Busenitz and Barney, 1997). In addition, it has a unique prediction, which can be tested in future empirical research: entrepreneurs in areas in which typical investors are individuals and area-specific funds would be more overconfident, than entrepreneurs in areas in which the typical investors are large multi-area funds or a government.

[20] Goel and Thakor (2008) also study how a risk-averse CEO's overconfidence enhances firm's value.

5.2 Overconfidence and Social Welfare

Our model can also explain how overconfidence can increase social welfare. Consider a society, where each agent i may either follow accepted guidelines or personal judgment when deciding how to work. This decision influences agent i's productivity $\mathbf{x}_i$, which may be either high or low. The payoff of each agent is a function of his output $\mathbf{x}_i$ and the total output $\sum_j \mathbf{x}_j$: $u_i = h\left(\mathbf{x}_i, \sum_j \mathbf{x}_j\right)$. Function h is assumed to be strictly increasing and concave in both parameters. For example, this is the case if a fixed amount of each agent's output is taxed and is being used to produce a public good. Alternatively, it might be that the output of each agent has a direct positive externality on other agents.

Calibrated agents (without confidence-bias) would follow the public signal too often, and obtain an inefficient outcome, in which the variance of the total productivity $\sum_j \mathbf{x}_j$ is too high. A utilitarian social planner would act as if it were a risk-averse principal in our model. Such a planner would like to induce social norms in favor of moderate overconfidence. This may explain why casual observation suggests that there are social norms in favor of moderate overconfidence (e.g., "self trust is the first secret of success", Ralph Waldo Emerson, 1803-1882).

6 Variants and Extensions

6.1 Choosing the Number of Agents

In the basic model we assumed that the number of agents is large. In this subsection we relax this assumption. Specifically, we allow the principal to choose the number of agents he employs, and we show that it is optimal for the principal to hire a large number of agents.

Proposition 8 shows that the principal strictly prefers to hire $k \cdot n$ agents than n agents.

Proposition 8 *For each $n \geq 1$ and $k \geq 2$ the principal can induce a strictly better outcome when the number of agents is $k \cdot n$ than when it is n.*

The intuition of Proposition 8 is that having more agents enables the principal to achieve better diversification. Each bias profile $(g_i)_{i \in I}$ with n agents can be replaced by a similar profile with $k \cdot n$ agents, in which each bias function g_i is induced by k agents. It can be shown that the random number of successes in the game with $k \cdot n$ agents second-order stochastically dominates the number of successes in the game with n agents, and thus it is preferred by the principal.

The following example shows that increasing the number of agents (but not multiplying it) may be bad for the principal.

Example 9 *(Example 5 revisited) Let* $L = 0$, $H = 1$, $\rho = 1$, $f_{\mathbf{p}} \sim$ *uniform*$(0, 0.5)$, $\mathbf{q} = 0.7$ *and let the principal's utility* $h(x)$ *be* $2x$ *if* $x < 0.5$ *and* 1 *if* $x \geq 0.5$. *Recall that when there are two agents the principal can achieve payoff 0.775 by using an asymmetric bias profile: one agent always follows the accepted guidelines while the other agent always follows his personal judgment. When there are three agents, the principal's best payoff is only 0.75, and it is achieved by having two agents always follow the accepted guidelines, and one agent always follows his personal judgment. The intuition why 3 agents are worse than 2 agents is that, the definition of utility h implies that the principal mainly cares that at least half of his agents succeed. It is easier to achieve this objective when there are only 2 agents (1 of them should succeed) rather then when there are 3 agents (and 2 of them should succeed).*

We can use Proposition 8 to demonstrate that our results do not depend on the assumption that there is a single principal. Consider a setup where there are several risk-averse principals and many agents, and that there is a small marginal cost for each additional hired agent. Due to the risk aversion of the principals and Proposition 8, each principal would choose to hire many agents, and all principals would prefer to hire overconfident agents.

6.2 Bias With Respect to Following the Accepted Guidelines

In the basic model we assume that agents can only have confidence bias with respect to their personal judgment, but not with respect to the accepted guidelines. In this subsection, we observe that this assumption is without loss of generality.

Consider a more general model, where the bias of each agent i is described by two functions $(g_{i,1}, g_{i,2})$ from $[0,1]$ to $[0,1]$, where $g_{i,1}$ is the bias with respect to the personal judgment (accuracy p_i is perceived as $g_{i,1}(p_i)$) and $g_{i,2}$ is the bias with respect to following the accepted guidelines (accuracy q is perceived by agent i as $g_{i,2}(q)$). Observe that the choice of agent i between the two actions only depends on the composite function $(g_{i,2})^{-1} \circ g_{i,1}$. This is because agent i chooses to follow the accepted guidelines if $g_{i,1}(\mathbf{p}_i) < g_{i,2}(\mathbf{q}) \Leftrightarrow (g_{i,2})^{-1} \circ g_{i,1}(\mathbf{p}_i) < \mathbf{q}$. This implies that our results remain the same in this extension. In particular, the optimal profile is such that each agent i has bias functions $(g_{i,1}, g_{i,2})$ that satisfy $(g_{i,2})^{-1} \circ g_{i,1} = g^*$, where g^* satisfies all the properties that were characterized in Theorems 3-6.

Thus one can interpret g^* as the excess bias in estimating the success probability of personal judgment relative to the bias in evaluating the success probability of accepted guidelines. Such an excess confidence bias is related to experimental stylized facts on overconfidence (discussed in Section 2): (1) when following personal judgment most of the uncertainty is internal, and this induces more overconfidence; and (2) it seems plausible to assume that the evaluation of the success probability of the accepted guidelines is based on many "weak" pieces of information: successes and failures of these guidelines in related decisions of different agents in the past; experimental evidence (such as the

strength-weight effect and set-based evaluations) suggests that such evaluations induce less overconfidence.

6.3 Costly Private Signals

The basic model assumes that private signals are costless. In this subsection we relax this assumption and extend our results to a more general framework that allows private signals to have cost. In the extended model, an independent random variable $0 \leq \mathbf{t}_i \sim f_{\mathbf{t}} \leq 1$ is assigned to each agent $i \in N$. Variable $\mathbf{t}_i$ is interpreted as the effectiveness of agent i in improving the accuracy of his personal judgment.

After agents are publicly informed about the value of $\mathbf{q}$ (the accuracy of following the accepted guidelines), each agent is privately informed of $\mathbf{t}_i$. Then, each agent privately chooses an effort level $0 \leq e_i \leq 1$, and receives private signal $p_i = p(e_i, \mathbf{t}_i)$ - the success probability of following his own personal judgment, where p is a strictly increasing function (in both parameters), and it is strictly concave in the effort level e_i. The payoff of each agent is either H (success) or L (failure) minus a cost of $(H - L) \cdot e_i$ for investing effort e_i. The rest of the model is the same as the basic model.

Let $p_{\mathbf{t}_i} \in [0, 1]$ be the unique number that maximizes $p(e_i, \mathbf{t}_i) - e_i$ (uniqueness holds due to concavity). The distribution of effectiveness levels $f_{\mathbf{t}}$ induces a unique distribution of maximizing accuracy levels $f_{p_{\mathbf{t}}}$. The following proposition asserts that our results also hold in this extended model, where $f_{p_{\mathbf{t}}}$ replaces $f_{\mathbf{p}}$.

Proposition 10 *The extended model with costly signals admits a unique optimal bias function g^*, which is the same as the optimal bias function g^* of the basic model with $f_{\mathbf{p}} = f_{p_{\mathbf{t}}}$.*

6.4 Agents as Experts

Consider a variant of the basic model in which at stage 2 each agent recommends an action (follow accepted guidelines or personal judgment), and the principal chooses the profile of actions $(a_i)_{i \in N}$ based on these recommendations. That is, each agent i is an informed expert, who advises the principal what to do in his area of expertise. Each expert's payoff remains the same: high payoff if the recommended action is successful, and low payoff otherwise.

If all agents are calibrated ($g(p) = p$), then too many of them would recommend the principal to follow the accepted guidelines (all experts i with $\mathbf{p}_i < \mathbf{q}$). The principal can gain higher payoff relative to the basic model, by violating some of these recommendations. However, his inability to separate agents with inaccurate private signals ($\mathbf{p}_i$ is substantially smaller than $\mathbf{q}$) from agents with relative accurate private signals limits his payoff.

Observe that this variant yields the same optimal bias function g^* as the basic model. This is because agents that follow g^* induce the principal's first-best payoff. Such agents behave as if they have the same utility as the principal including his interest in diversification. Thus, the principal will always choose to follow the recommendations of such g^*-biased experts.

6.5 Risk-Averse Agents

In the basic model the utility of each agent is equal to

$$\mathbf{u}_i\left(a_{guidelines}\right)=\begin{cases}H & \text{if } \xi_{i,q}=1,\\ L & \text{if } \xi_{i,q}=0,\end{cases} \text{ and } u_i\left(a_{original}\right)=\begin{cases}H & \text{if } \xi_{i,p}=1,\\ L & \text{if } \xi_{i,p}=0,\end{cases}$$

and the utility of the principal is a concave function of the average utility of the agents. Thus, in the basic model the principal is more risk-averse than the agents (for example, when there is a single agent, the principal's utility is a concave function of the agent's utility). This may seem implausible in some applications.

However, this assumption can be relaxed without changing the results as follows (using the fact that each agent faces only two possible outcomes). We reinterpret $\mathbf{u}_i$ as a monetary payoff, and we allow the utility of agent i to be any monotone function of this monetary payoff: $h_i\left(\mathbf{u}_i\right)$. Specifically, our results (Theorem 3-7) also hold if each agent has utility function $h_i\left(x\right)$ that is more concave then the principal's utility $h\left(x\right)$.

6.6 Modeling Overconfidence as Underestimating Variance

We modeled overconfidence as overestimating the accuracy of discrete private signals. Another common way to model overconfidence, especially in finance models (e.g., Odean, 1998), is underestimating the variance of continuous private signals. In this subsection, we briefly demonstrate how our model can be reformulated to represent overconfidence in this way. For brevity, we only sketch the main details of a simple case that is analogous to the perfect correlation case ($\rho=1$).

Let the random variable $0<\sigma_q\sim f_q$ be the variance of the public signal, and for each $i\in I$ let the random variable $0<\sigma_{p_i}\sim f_p$ be the variance of the private signal of agent i (where f_q and f_p are continuous distributions, and the variables $\left(\sigma_q,\left(\sigma_{p_i}\right)_{i\in I}\right)$ are independent). All agents publicly receive σ_q. Let $\mathbf{R}_q\sim norm\left(0,\sigma_q\right)$ and for each $i\in I$ let $\mathbf{R}_{p_i}\sim norm\left(0,\sigma_{p_i}\right)$. At the first stage of the interaction the principal chooses a bias profile for the agents: $\left(g_i\right)_{i\in I}$. Each function $g_i:\mathbb{R}^+\to\mathbb{R}^+$ describes the bias function of agent i relative to the private signal. That is, agent i is privately misinformed that the value of σ_{p_i} is $g_i\left(\sigma_{p_i}\right)$. At the second stage of the interaction, each agent chooses one of two actions: $\{a_{guidelines},a_{original}\}$. If agent i chooses $a_{guidelines}$ ($a_{original}$) then his payoff is

$-\mathbf{R}_q^2\left(-\mathbf{R}_{p_i}^2\right)$. The payoff of the principal is a concave increasing function of the average payoff of the agents. Similar to Theorem 3, one can show that there is a unique first-best homogeneous optimal bias profile which represents overconfidence: $g^*(\sigma_{p_i}) < \sigma_{p_i}$.

A Proofs

A.1 Preliminaries

The following lemma presents an equivalent formulation for the decreasing absolute risk aversion property, which will be used in the proofs of Theorem 3 (part 3) and Theorem 6.

Lemma 11 Let $h(y)$ be a strictly concave increasing function. Then $h(y)$ satisfies (strictly) decreasing absolute risk aversion (DARA) if and only if $f_a(y) = \frac{h'(y)}{h'(y+a)}$ is a strictly decreasing function of y for each $a > 0$.

PROOF. The lemma is proven as follows:

$h(y)$ satisfies DARA $\Leftrightarrow$ for every $y, a > 0$:

$$r_A(y) > r_A(y+a) \Leftrightarrow -\frac{h''(y)}{h'(y)} > -\frac{h''(y+a)}{h'(y+a)}$$

$$\Leftrightarrow \frac{h''(y)}{h'(y)} < \frac{h''(y+a)}{h'(y+a)} \Leftrightarrow h''(y) \cdot h'(y+a) - h''(y+a) \cdot h'(y) < 0$$

$$\Leftrightarrow f_a'(y) = \frac{h''(y) \cdot h'(y+a) - h''(y+a) \cdot h'(y)}{(h'(y+a))^2} < 0$$

$\Leftrightarrow f_a(y)$ is strictly decreasing. $\square$

A.2 Proof of Theorem 3

Theorem 3 There exists a unique optimal bias function g^*, which induces the first-best payoff, with the following properties:

(1) $g^*(p) > p$ for every $0 < p < 1$ (overconfidence).
(2) g^* is continuous.
(3) g^* is increasing: $\frac{dg^*(p)}{dp} > 0$ for every $0 < p < 1$, $g^*(0) = 0$, and $g^*(1) = 1$.
(4) g^* does not depend on the distribution f_q.

PROOF. The proof includes two parts. The first part shows that the first-best outcome of the principal can be approximately induced by a bias function. The second part characterizes this optimal bias function g^*, and shows its uniqueness.

Approximating the first-best payoff by a bias function

We begin by dealing with the "first-best" case in which the principal receives all the private signals $(\mathbf{p}_i)_{i\in I}$ and the public signal $\mathbf{q}$ and chooses the actions of all the agents. Without loss of generality the first-best strategy is a function ϕ that chooses a threshold $p = \phi(q, p_1, ..., p_n)$, such that each agent i with higher (lower) accuracy level $p_i \geq p$ ($p_i < p$) follows his personal judgment (accepted guidelines). The expected payoff (u) of this threshold is:

$$\mathbf{E}\left(h\left(L+(H-L)\left(\frac{1}{n}\left(\#\left\{i|p_i<p \text{ and } \xi_{i,q}=1\right\}+\#\left\{i|p_i\geq p \text{ and } \xi_{i,p}=1\right\}\right)\right)\right)|q,(p_i)_{i\in I}\right).$$

Variables $(\xi_{i,q}, \xi_{i,p})_{i\in I}$ are conditionally independent given ξ_q. Assuming that the number of agents is large enough, the expected payoff is well approximated by

$$u = P(\xi_q=1)\cdot h\left(L+\frac{H-L}{n}\cdot\left(\sum_{p_i<p}P(\xi_{i,q}=1|\xi_q=1)+\sum_{p_i\geq p}P(\xi_{i,p}=1)\right)\right)+$$
$$P(\xi_q=0)\cdot h\left(L+\frac{H-L}{n}\cdot\left(\sum_{p_i<p}P(\xi_{i,q}=1|\xi_q=0)+\sum_{p_i\geq p}P(\xi_{i,p}=1)\right)\right)+o(\epsilon).$$

Substituting the different probabilities yields the following:

$$u = q\cdot h\left(L+\frac{H-L}{n}\cdot\left((\sqrt{\rho}+(1-\sqrt{\rho})\cdot q)\cdot\#\left\{i|p_i<p\right\}+\sum_{p_i\geq p}p_i\right)\right)+$$
$$(1-q)\cdot h\left(L+\frac{H-L}{n}\cdot\left(((1-\sqrt{\rho})\cdot q)\cdot\#\left\{i|p_i<p\right\}+\sum_{p_i\geq p}p_i\right)\right)+o(\epsilon).$$

To simplify notation let $f = f_p$ and $F = F_p$. Assuming again that the number of agents is large enough, one can approximate the empirical distribution of the private signals $(p_1, ..., p_n)$ by their prior distribution f. This gives the following approximation:

$$u = q \cdot h\left(L + (H-L) \cdot \left(\left(\sqrt{\rho} + (1-\sqrt{\rho}) \cdot q\right) \cdot F(p) + \int_p^1 x \cdot f(x)\, dx\right)\right) + \quad \text{(A.1)}$$
$$(1-q) \cdot h\left(L + (H-L) \cdot \left(\left((1-\sqrt{\rho}) \cdot q\right) \cdot F(p) + \int_p^1 x \cdot f(x)\, dx\right)\right) + o(\epsilon).$$

Consider the bias function $g^*(p)$ that is defined as follows: $p = (g^*)^{-1}(q)$ is the threshold that maximizes Eq. A.1 (neglecting the error term $o(\epsilon)$). By the above arguments, such a bias function ϵ-induces the first-best payoff.

Characterizing the unique optimal bias function $g^*(p)$

We now calculate the value of $p = (g^*)^{-1}(q)$ that maximizes Eq. A.1 (neglecting the error term $o(\epsilon)$). One can verify that $g^*(0) = 0$ and $g^*(1) = 1$. We focus on the case $0 < q < 1$. Observe first that the optimal p must be in the interval $\left(\left(1-\sqrt{\rho}\right) \cdot q, q\right)$ because: (1) following $a_{guidelines}$ strictly dominates following $a_{original}$ when $p_i \leq \left(1-\sqrt{\rho}\right) \cdot q$, as the former is better than the latter even in the "bad" state of nature (in which $\xi_q = 0$); and (2) following $a_{original}$ strictly dominates $a_{guidelines}$ when $p_i > q$ as it yields a 2nd-order stochastic dominant payoff.

To simplify notation let:

$$A_{p,q,\rho} = \left((1-\sqrt{\rho}) \cdot q\right) \cdot F(p) + \int_p^1 x \cdot f(x)\, dx.$$

Observe the following properties of $A_{p,q,\rho}$:

(1) $A_{p,q,\rho}$ is strictly decreasing in p in the interval $\left(\left(1-\sqrt{\rho}\right) \cdot q, q\right)$ (because $\frac{dA_{p,q,\rho}}{dp} = \left(\left(1-\sqrt{\rho}\right) \cdot q - p\right) \cdot f(p)$ is negative for every $p > \left(1-\sqrt{\rho}\right) \cdot q$).
(2) $A_{p,q,\rho} + \sqrt{\rho} \cdot F(p)$ is strictly increasing in p in the interval $\left(\left(1-\sqrt{\rho}\right) \cdot q, q\right)$ (because

$$\frac{d\left(A_{p,q,\rho} + \sqrt{\rho} \cdot F(p)\right)}{dp} = \left((1-\sqrt{\rho}) \cdot q + \sqrt{\rho} - p\right) \cdot f(p) > \left((1-\sqrt{\rho}) \cdot q + \sqrt{\rho} \cdot q - p\right) \cdot f(p)$$

is positive for every $p < q$.
(3) $A_{p,q,\rho}$ is weakly increasing in q (strictly increasing when $\rho < 1$).
(4) $A_{p,q,\rho}$ is strictly decreasing in ρ.
(5) $A_{p,q,\rho} + \sqrt{\rho} \cdot F(p)$ is strictly increasing in ρ.

For every $0 < q < 1$ we find $p = (g^*)^{-1}(q)$ by derivation:

$$\frac{du}{dp} = q \cdot h'\left(L + (H-L) \cdot \left(\sqrt{\rho} \cdot F(p) + A_{p,q,\rho}\right)\right)\left(\left((1-\sqrt{\rho}) \cdot q + \sqrt{\rho} - p\right) f(p)\right)(H-L)$$
$$+ (1-q) \cdot h'\left(L + (H-L) \cdot A_{p,q,\rho}\right)\left(\left((1-\sqrt{\rho}) \cdot q - p\right) f(p)\right)(H-L).$$

Assuming an internal solution ($\frac{du}{dp}=0$) yields:

$$\frac{h'\left(L+(H-L)\cdot A_{p,q,\rho}\right)}{h'\left(L+(H-L)\cdot\left(A_{p,q,\rho}+\sqrt{\rho}\cdot F\left(p\right)\right)\right)}=\frac{q\cdot\left(\sqrt{\rho}+\left(1-\sqrt{\rho}\right)\cdot q-p\right)}{(1-q)\cdot\left(p-\left(1-\sqrt{\rho}\right)\cdot q\right)}. \tag{A.2}$$

Using the strict concavity of h, the fact that $A_{p,q,\rho}$ is strictly decreasing in p and $A_{p,q,\rho}+\sqrt{\rho}\cdot F\left(p\right)$ is strictly increasing in p in the interval $\left(\left(1-\sqrt{\rho}\right)\cdot q,q\right)$ implies that the left-hand side (l.h.s.) of Eq. A.2 is a strictly increasing function of p. Observe that the right-hand side (r.h.s.) is a strictly decreasing function of p, and that for p close enough to $\left(1-\sqrt{\rho}\right)\cdot q$ the r.h.s. is larger than the l.h.s. (as the r.h.s. converges to ∞ when p converges to $\left(1-\sqrt{\rho}\right)\cdot q$), while for p close enough to q the l.h.s. is larger than the r.h.s. (as the r.h.s. converges to 1 when p converges to q while the l.h.s. is always larger than 1). Thus for each $0<q<1$ there is a unique solution $p=\left(g^*\right)^{-1}\left(q\right)$ to Eq. A.2 in the interval $\left(\left(1-\sqrt{\rho}\right)\cdot q,q\right)$, which is a continuous function of q. In particular, this implies that $\left(g^*\right)^{-1}\left(q\right)<q$ for every $0<q<1$ (the overconfidence property).

By Lemma 11, the l.h.s. of Eq. A.2 is weakly decreasing in q (as $A_{p,q,\rho}$ is weakly increasing in q). One can verify that the r.h.s. is strictly increasing in q. Thus increasing q by $\delta>0$ while holding p constant, would make the r.h.s. larger than the l.h.s., and p must be increased in order to retain the equality in Eq. A.2. This implies that $\left(g^*\right)^{-1}\left(q+\delta\right)>\left(g^*\right)^{-1}\left(q\right)$ for every $0<q<1$ and $1-q>\delta>0$, and because of this, $\left(g^*\right)^{-1}\left(q\right)$ is a strictly increasing function of q.

One can verify that $\frac{du}{dp}>0$ for every $p<\left(g^*\right)^{-1}\left(q\right)$, and $\frac{du}{dp}<0$ for every $p>\left(g^*\right)^{-1}\left(q\right)$. Thus, any other bias threshold $p\neq\left(g^*\right)^{-1}\left(q\right)$ would yield a strictly lower expected payoff.

The above arguments show that the profile in which all agents have bias g^* induces (up to ϵ) the first-best outcome for the principal (and thus it is ϵ-optimal), that g^* has all the required properties (overconfidence, continuous, and increasing), and that g^* is unique in the following sense: any other bias function $\tilde{g}$ such that $\tilde{g}\neq g^*$ on a set with a positive Lebesgue measure yields a strictly lower payoff, assuming the number of agents is sufficiently large. Observe (Eq. A.2), that $g^*\left(p\right)$ does not depend on the distribution $f_{\mathbf{q}}$. □

A.3 Proof of Theorem 4

Theorem 4 Let $(g_i)_{i\in I}$ be an heterogeneous profile. Then there is $k_0\in\mathbb{N}$ such that for every $k\geq k_0$, there is an homogeneous profile that induces a strictly better outcome than $(g_i)_{i\in I}$ in the game with $k\cdot|I|$ agents.

PROOF. Let $\tilde{g}$ be the following bias function (homogeneous bias profile): for each $q\in[0,1]$, $\left(\tilde{g}\right)\left(q\right)^{-1}$ is the unique solution to the following equation:

$$F\left(\left(\tilde{g}\right)^{-1}(q)\right)=\sum_{i\in I}\frac{1}{|I|}\left(F\left(g_i^{-1}(q)\right)\right)$$

That is, $\tilde{g}$ is a bias function that averages the heterogeneous profile $(g_i)_{i\in I}$. Fix any $0<q<1$ satisfying $\min_i\left(g_i\right)^{-1}(q)<\max_i\left(g_i\right)^{-1}(q)$. To simplify notation let $p_i=g_i^{-1}(q)$ and $\tilde{p}=\tilde{g}^{-1}(q)$. By the arguments given in the previous subsection, for each q, the expected payoff of the heterogeneous profile $(g_i)_{i\in N}$ in the game with $k\cdot|I|$ agents (for large enough k) is approximately given by

$$q\cdot h\left(L+(H-L)\cdot\left(\frac{1}{|I|}\sum_{i\in I}\left(\left(\sqrt{\rho}+(1-\sqrt{\rho})\cdot q\right)\cdot F(p_i)+\int_{p_i}^{1}x\cdot f(x)\,dx\right)\right)\right)+$$
$$(1-q)\cdot h\left(L+(H-L)\cdot\left(\frac{1}{|I|}\sum_{i\in I}\left(\left((1-\sqrt{\rho})\cdot q\right)\cdot F(p_i)+\int_{p_i}^{1}x\cdot f(x)\,dx\right)\right)\right),$$

and the expected payoff of the homogeneous profile $\tilde{g}$ is approximately given by

$$q\cdot h\left(L+(H-L)\cdot\left(\left(\sqrt{\rho}+(1-\sqrt{\rho})\cdot q\right)\cdot F(\tilde{p})+\int_{\tilde{p}}^{1}x\cdot f(x)\,dx\right)\right)+$$
$$(1-q)\cdot h\left(L+(H-L)\cdot\left(\left((1-\sqrt{\rho})\cdot q\right)\cdot F(\tilde{p})+\int_{\tilde{p}}^{1}x\cdot f(x)\,dx\right)\right).$$

As $F(\tilde{p})=\sum_{i\in I}\frac{1}{|I|}\left(F(\tilde{p}_i)\right)$, the homogeneous profile yields a higher expected payoff if and only if

$$\frac{1}{|I|}\sum_{i\in I}\int_{p_i}^{1}xf(x)\,dx<\int_{\tilde{p}}^{1}xf(x)\,dx.$$

This is equivalent to

$$\frac{1}{|I|}\sum_{i\in I}\left(\int_{p_i}^{1}xf(x)\,dx-\int_{\tilde{p}}^{1}xf(x)\,dx\right)<0,$$

or equivalently (using the notation that $\int_a^b f(x)\,dx=-\int_b^a f(x)\,dx$ when $b<a$):

$$\frac{1}{|I|}\sum_{i\in I}\int_{p_i}^{\tilde{p}}xf(x)\,dx<0,$$

which holds if and only if

$$\frac{1}{n}\sum_{i\in I}\left(F\left(\tilde{p}\right)-F\left(p_i\right)\right)\cdot\mathbf{E}\left(\mathbf{p}|\min\left(p_i,\tilde{p}\right)\leq\mathbf{p}\leq\max\left(p_i,\tilde{p}\right)\right)<0. \tag{A.3}$$

Observe that

$$\frac{1}{n}\sum_{i\in I}\left(F\left(\tilde{p}\right)-F\left(p_i\right)\right)=0,$$

and that $\mathbf{E}\left(\mathbf{p}|\min\left(p_i,\tilde{p}\right)\leq\mathbf{p}\leq\max\left(p_i,\tilde{p}\right)\right)$ is strictly increasing in p_i and strictly decreasing in $\left(F\left(\tilde{p}\right)-F\left(p_i\right)\right)$. This implies that Inequality A.3 holds.

The above arguments show that for each q such that $\min_i\left(g_i\right)^{-1}\left(q\right)<\max_i\left(g_i\right)^{-1}\left(q\right)$, $\tilde{g}$ has higher expected value than $\left(g_i\right)_{i\in I}$, conditional on $\mathbf{q}=q$. The fact that $\left(g_i\right)_{i\in I}$ is a heterogeneous bias profile (i.e., that $\min_i\left(g_i\right)^{-1}\left(q\right)<\max_i\left(g_i\right)^{-1}\left(q\right)$ in a set with positive Lebesgue measure), implies that $\tilde{g}$ has higher expected value than $\left(g_i\right)_{i\in I}$ (without conditioning on the value of $\mathbf{q}$). By the law of large numbers, if the number of agents is sufficiently large then it implies that with high probability $\tilde{g}$ induces a strictly larger payoff than $\left(g_i\right)_{i\in I}$. □

A.4 Proof of Theorem 6

Theorem 6 Let $\left(I_1,\rho_1,f_{\mathbf{q},1},f_{\mathbf{p},\mathbf{1}},L_1,H_1,h_1\right)$ and $\left(I_2,\rho_2,f_{\mathbf{q},2},f_{\mathbf{p},2},L_2,H_2,h_2\right)$ be two sets of parameters of our model, and let $g_1^*\left(g_2^*\right)$ be the unique optimal bias function given the first (second) set of parameters. Then g_1^* presents more overconfidence ($g_1^*\left(p\right)>g_2^*\left(p\right)$ for every $0<p<1$) in each of the following cases:

(1) Utility h_1 is more risk-averse than h_2 and all other parameters are the same. That is, $h_1=\psi\circ h_2$ where ψ is concave and increasing.
(2) The first correlation coefficient is larger ($\rho_1>\rho_2$), and all other parameters are the same.

PROOF.

(1) Let $h_2=h$ and let $h_1=\psi\circ h$ where ψ is concave and increasing. Substituting $\psi\circ h$ as the principal's utility in Eq. (A.2) yields the following equation:

$$\frac{\Psi'\left(h\left(L+\left(H-L\right)\cdot A_{p,q,\rho}\right)\right)\cdot h'\left(L+\left(H-L\right)\cdot A_{p,q,\rho}\right)}{\Psi'\left(h\left(L+\left(H-L\right)\cdot\left(A_{p,q,\rho}+\sqrt{\rho}\cdot F\left(p\right)\right)\right)\right)h'\left(L+\left(H-L\right)\cdot\left(A_{p,q,\rho}+\sqrt{\rho}\cdot F\left(p\right)\right)\right)}$$
$$=\frac{q\cdot\left(\sqrt{\rho}+\left(1-\sqrt{\rho}\right)\cdot q-p\right)}{\left(1-q\right)\cdot\left(p-\left(1-\sqrt{\rho}\right)\cdot q\right)}. \tag{A.4}$$

Let $\left(p,q_2\right)$ be any solution to Eq. (A.2): $q_2=g_2^*\left(p\right)$. We now substitute $\left(p,q_2\right)$ in Eq. A.4. Observe that the l.h.s. of Eq. (A.4) is larger than the l.h.s. of Eq. (A.2) due to the concavity of ψ, while the r.h.s. of both equations are the same. This implies that

with the values of (p, q_2), the l.h.s. is larger then the the r.h.s. of Eq. (A.4). Recall that the l.h.s is weakly decreasing in q (due to Lemma 11) while the r.h.s. is strictly increasing in q. This implies that $q_1 = g_1^*(p) > q_2$.

(2) Let (p, q_2) be any solution to Eq. (A.2) given ρ_2: $q_2 = g_2^*(p)$. Observe that: (1) The r.h.s. of Eq. (A.2) is strictly decreasing in ρ, and (2) the concavity of h implies that the l.h.s. of Eq. (A.2) is strictly increasing in ρ (recall that $A_{p,q,\rho}$ is strictly decreasing in ρ, while $A_{p,q,\rho} + \sqrt{\rho} \cdot F(p)$ is strictly increasing in ρ). Thus, given ρ_1, p and q_2, the l.h.s. is strictly larger than the r.h.s. As the l.h.s is decreasing in q and the r.h.s. is strictly increasing in q, it implies that: $q_1 = g_1^*(p) > q_2$. □

A.5 Proof of Theorem 7

Theorem 7 Assume that the principal has a CRRA utility function. Then,

(1) g^* depends on the payoffs only through its dependency on the potential gain $D = (H - L)/L$.

(2) If $D_1 > D_2$ (and all other parameters are the same) then g_1^* presents more overconfidence ($g_1^*(p) > g_2^*(p)$ for every $0 < p < 1$).

(3) If distribution $f_{\mathbf{p},2}$ has a first order stochastic dominance over $f_{\mathbf{p},\mathbf{1}}$ (and all other parameters are the same) then g_1^* presents more overconfidence.

(4) The ratio between the perceived error probability and the true error probability of personal judgment ($\frac{1-g(p)}{1-p}$) converges to $(1/(D+1))^{\phi} = \left(\frac{L}{H}\right)^{\phi}$ when both p and ρ converge to 1.

PROOF.

(1) Placing $h'(x) = x^{-\phi}$ in Eq. (A.2) yields:

$$\frac{(L + (H-L) \cdot A_{p,q,\rho})^{-\phi}}{\left(L + (H-L) \cdot \left(A_{p,q,\rho} + \sqrt{\rho} \cdot F(p)\right)\right)^{-\phi}} = \frac{q \cdot \left(\left(\sqrt{\rho} + \left(1 - \sqrt{\rho}\right) \cdot q\right) - p\right)}{(1-q) \cdot \left(p - \left(\left(1 - \sqrt{\rho}\right) \cdot q\right)\right)} \Rightarrow$$

$$\left(\frac{\left(L + (H-L) \cdot \left(A_{p,q,\rho} + \sqrt{\rho} \cdot F(p)\right)\right)}{L + (H-L) \cdot A_{p,q,\rho}}\right)^{\phi} = \frac{q \cdot \left(\left(\sqrt{\rho} + \left(1 - \sqrt{\rho}\right) \cdot q\right) - p\right)}{(1-q) \cdot \left(p - \left(\left(1 - \sqrt{\rho}\right) \cdot q\right)\right)} \Rightarrow$$

$$\left(\frac{\left(\frac{L}{L} + \frac{H-L}{L} \cdot \left(A_{p,q,\rho} + \sqrt{\rho} \cdot F(p)\right)\right)}{\frac{L}{L} + \frac{H-L}{L} \cdot A_{p,q,\rho}}\right)^{\phi} = \frac{q \cdot \left(\left(\sqrt{\rho} + \left(1 - \sqrt{\rho}\right) \cdot q\right) - p\right)}{(1-q) \cdot \left(p - \left(\left(1 - \sqrt{\rho}\right) \cdot q\right)\right)} \Rightarrow$$

$$\left(1 + \frac{\frac{H-L}{L} \cdot \sqrt{\rho} \cdot F(p)}{1 + \frac{H-L}{L} \cdot A_{p,q,\rho}}\right)^{\phi} = \frac{q \cdot \left(\left(\sqrt{\rho} + \left(1 - \sqrt{\rho}\right) \cdot q\right) - p\right)}{(1-q) \cdot \left(p - \left(\left(1 - \sqrt{\rho}\right) \cdot q\right)\right)}.$$

Substituting $D = \frac{H-L}{L}$ and $A_{p,q,\rho} = \left(\left(1-\sqrt{\rho}\right)\cdot q\right)\cdot F(p) + \int_p^1 x \cdot f(x)\,dx$ gives

$$\left(1 + \frac{D\cdot\sqrt{\rho}\cdot F(p)}{1+D\left(\left(\left(1-\sqrt{\rho}\right)\cdot q\right)\cdot F(p) + \int_p^1 x\cdot f(x)\,dx\right)}\right)^{\phi} = \frac{q\cdot\left(\left(\sqrt{\rho}+\left(1-\sqrt{\rho}\right)\cdot q\right)-p\right)}{(1-q)\cdot\left(p-\left(\left(1-\sqrt{\rho}\right)\cdot q\right)\right)}. \tag{A.5}$$

This proves that the g^* depends on the payoffs only through its dependence on D.

(2) Observe that the l.h.s. of Eq. (A.5) increases in D. Similar to the arguments in the proof of Theorem 6, this implies that larger D induces more overconfidence.

(3) Let $f_{\mathbf{p},2}$ be a distribution with first order stochastic dominance over $f_{\mathbf{p},\mathbf{1}}$. That is, $F_2(p) < F_1(p)$ for every $0 < p < 1$. We have to show that $g_1^* \geq g_2^*$. Observe that the l.h.s. of Eq. (A.5) is larger when $f_{\mathbf{p},1}$ replaces $f_{\mathbf{p},2}$. This is because

$$\left(1 + \frac{D\cdot\sqrt{\rho}\cdot F_2(p)}{1+D\left(\left(\left(1-\sqrt{\rho}\right)\cdot q\right)\cdot F_2(p) + \int_p^1 x\cdot f(x)\,dx\right)}\right)^{\phi} =$$

$$\left(1 + \frac{D\cdot\sqrt{\rho}\cdot F_2(p)}{1+D\left(\left(\left(1-\sqrt{\rho}\right)\cdot q\right)\cdot F_2(p) + \int_p^1 (1-F_2(x))\,dx\right)}\right)^{\phi} =$$

$$\left(1 + \left(\frac{1+D\left(\left(\left(1-\sqrt{\rho}\right)\cdot q\right)\cdot F_2(p) + \int_p^1 (1-F_2(x))\,dx\right)}{D\cdot\sqrt{\rho}\cdot F_2(p)}\right)^{-1}\right)^{\phi} =$$

$$\left(1 + \left(\frac{D\left(\left(1-\sqrt{\rho}\right)\cdot q\right)\cdot F_2(p)}{D\cdot\sqrt{\rho}\cdot F_2(p)} + \frac{1+D\int_p^1 (1-F_2(x))\,dx}{D\cdot\sqrt{\rho}\cdot F_2(p)}\right)^{-1}\right)^{\phi} \leq$$

$$\left(1 + \left(\frac{\left(\left(1-\sqrt{\rho}\right)\cdot q\right)}{\sqrt{\rho}} + \frac{1+D\int_p^1 (1-F_2(x))\,dx}{D\cdot\sqrt{\rho}\cdot F_1(p)}\right)^{-1}\right)^{\phi} \leq$$

$$\left(1 + \left(\frac{\left(\left(1-\sqrt{\rho}\right)\cdot q\right)}{\sqrt{\rho}} + \frac{1+D\int_p^1 (1-F_1(x))\,dx}{D\cdot\sqrt{\rho}\cdot F_1(p)}\right)^{-1}\right)^{\phi}.$$

Similar to the arguments in the proof of Theorem 6, $g_1^*(p) \geq g_2^*(p)$ for every p.

(4) Let both p and ρ converge to 1 (which implies that q also converges to 1). Substituting it in Eq. (A.5) yields (approximately)

$$\left(\frac{H}{L}\right)^{\phi} = (1+D)^{\phi} \approx \frac{(1-p)}{(1-q)},$$

and that completes the proof. □

A.6 Proof of Proposition 8

Proposition 8 For each $n \geq 1$ and $k \geq 2$ the principal can induce a strictly better outcome when the number of agents is $k \cdot n$ than when it is n.

PROOF. Let $(g_i)_{i \in I}$ be a bias profile in the game with $n = |I|$ agents. Recall that for each agent $i \in I$, $\mathbf{u}_i$ is the random payoff of agent i with bias function g_i, and that the principal's payoff is $h\left(\frac{1}{n}\sum_{i \in I} \mathbf{u}_i\right)$. Consider $(g_i)_{i \in I}$ as a profile in the game with $k \cdot n$ agents (where each k agents share one of the bias functions g_i). This profile induces the following payoff:

$$h\left(\frac{1}{n}\sum\nolimits_{i \in I}\frac{1}{k}\sum_{j=1}^{k}\mathbf{u}_{(i-1)\cdot k+j}\right),$$

where for each i, the variables $\left\{\left(\mathbf{u}_{(i-1)\cdot k+j}\right)_{j=1,\ldots,k}, u_i\right\}$ are identically distributed. Observe that $\frac{1}{n}\sum_{i \in I}\frac{1}{k}\sum_{j=1}^{k}\mathbf{u}_{(i-1)\cdot k+j}$ second-order stochastically strictly dominates $\frac{1}{n}\sum_{i \in I}\mathbf{u}_i$. By the concavity of h, it implies that the principal strictly prefers the outcome in the game with $k \cdot n$ agents. Thus, any outcome in the game with n agents is strictly dominated by an outcome in the game with $k \cdot n$ agents. □

A.7 Proof of Proposition 10

Proposition 10 The extended model with costly signals admits a unique optimal bias function g^*, which is the same as the optimal bias function g^* of the basic model with $f_{\mathbf{p}} = f_{p_{\mathbf{t}}}$.

PROOF. We begin by calculating the first-best profile in a game with many agents $n >> 1$. Without loss of generality for each $q \in [0,1]$, there is some effectiveness value $t_0 = \alpha(q)$ such that the optimal payoff can be induced by all agents using the same threshold strategy: (1) agents with low effectiveness ($\mathbf{t}_i < t_0$) do not invest any effort and follow the accepted guidelines, and (2) agents with high effectiveness ($\mathbf{t}_i \geq t_0$) invest some effort and follow personal judgment.

Consider an agent with high effectiveness: $\mathbf{t} \geq t_0$. His expected payoff from investing effort e is $L + (H - L) \cdot (p(e,t) - e)$. This is maximized in e_t^* that satisfies $\frac{d(p(e,t))}{de} = 1$ (a unique maximizer exists due to the strict concavity of $p(e,t)$). Let $p_t^* = p(e_t^*, t)$. For large enough n, if all agents with high effectiveness invest effort e_t^*, it ϵ-maximizes the principal's payoff (by the law of large numbers).

Let $p_0 = p_{t_0}^*$ be the success probability of an agent with threshold effectiveness value t_0. The choice of an optimal threshold t_0 is equivalent to the problem of finding the optimal

threshold p_0 in Theorem 3. Thus the unique optimal bias function g^* of the basic model (Section 3) is also optimal and unique in the extended model (with $f_{\mathbf{p}} = f_{p_\mathbf{t}}$). □

References

[1] **Alicke Mark D., Olesya Govorun.** 2005. "The better-than-average effect". In: *The self in social judgment*, ed. Mark D. Alicke, David A. Dunning, and Justin I. Krueger, NY: Psychology Press.

[2] **Barber, Brad M., and Terrance Odean.** 2001. "Boys will be boys: gender, overconfidence, and common stock investment". *The Quarterly Journal of Economics*, 116(1): 261-292.

[3] **Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey.** 2010. "Managerial Miscalibration" NBER Working Paper #w16215, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1648015.

[4] **Bénabou, Roland, and Jean Tirole**. 2002. "Self-confidence and personal motivation" *Quarterly Journal of Economics,* 117(3): 871-915.

[5] **Benoit, Jean P. and Juan Dubra**. 2011. "Apparent Overconfidence", *Econometrica,* forthcoming.

[6] **Blume, Lawrence, and David Easley**. 1992. "Evolution and Market Behavior" *Journal of Economic Theory*, 58, 9-40.

[7] **Bernardo, Antonio E., and Ivo Welch.** 2001. "On the evolution of overconfidence and entrepreneurs" *Journal of Economics & Management Strategy*, 10(3), 301-330.

[8] **Brenner, Lyle A., Derek J. Kohler, Varda Liberman, and Amos Tversky.** 1996. "Overconfidence in probability and frequency Judgments: a critical examination" *Organizational Behavior and Human Decision Processes,* 65(3): 212-219.

[9] **Budescu, David V., Ning Du**. 2007. "Coherence and consistency of investors' probability judgments" *Management Science*, 53(11): 1731-1744.

[10] **Budescu, David V., Thomas Wallsten, and Wing Tung Au**. 1997. "On the importance of random error in the study of probability judgment. Part II: applying the stochastic judgment model to detect systematic trends" *Journal of Behavioral Decision Making,* 10: 173-188.

[11] **Busenitz, Lowel W., Jay B. Barney**. 1997. "Differences between entrepreneurs and managers in large organizations: biases and heuristics in strategic decision-making" *Journal of Business Venturing*, 12: 9-30.

[12] **Chuang, Wen-I, and Bong-Soo Lee**. 2006. "An empirical evaluation of the overconfidence hypothesis" *Journal of Banking and Finance* 30, 2489-2515.

[13] **Compte, Oliver, and Andrew Postlewaite**. 2004. "Confidence-Enhanced Performance" *The American Economic Review*, 94(5), 1536-1557.

[14] **Cooper, Arnold C., Carolyn Y. Woo, and William C. Dunkelberg**. 1988. "Entrepreneurs' perceived chances for success" *Journal of Business Venturing* 3, 97-108.

[15] **Curry, Philip A.** 2001. "Decision making under uncertainty and the evolution of interdependent preferences" *Journal of Economic Theory,* 98: 357-369.

[16] **Erev, Ido, Thomas S. Wallsten, David V. Budescu**. 1994. "Simultaneous over and underconfidence: the role of error in judgment processes" *Psychological Review*, 101(3): 519-527.

[17] **Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein.** 1977. "Knowing with certainty: the appropriateness of extreme confidence" *Journal of Experimental Psychology: Human Perception and Performance,* 3(4): 552-564.

[18] **Friesen Geoffrey, Paul A Weller**. 2006. "Quantifying cognitive biases in analyst earnings forecasts" *Journal of Financial Markets*, 9: 333-365.

[19] **Gervais, Simon, J.B. Heaton, Terrance Odean**. 2010. "Overconfidence, Compensation Contracts, and Capital Budgeting" *Journal of Finance*, forthcoming.

[20] **Gervais, Simon, Terrance Odean**. 2001. "Learning to be overconfident" *The Review of Financial Studies*, 14(1): 1-27.

[21] **Gigerenzer, Gerd, Ulrich Hoffrage, and Hienz Kleinbolting.** 1991. "Probabilistic mental models: a Brunswikian theory of confidence" *Psychological Review*, 98(4): 506-528.

[22] **Glaser, Markus, Thomas Langer, and Martin Weber**. 2010. "Overconfidence of professionals and lay people: individual differences within and between tasks?" unpublished manuscript. http://ssrn.com/abstract=712583

[23] **Goel, Anand M, Anjan V. Thakor**. 2008. "Overconfidence, CEO selection and Corporate Governance" *The Journal of Finance*, LXIII(6): 2737-2783.

[24] **Griffin, Dale, and Lyle Brenner.** 2004. "Perspectives on probability judgment calibration" In: *Blackwell Handbook of Judgment and Decision Making*, ed. Derek J. Koehler and Nigel Harvey, 177–99, Boston: Blackwell Publishing.

[25] **Griffin, Dale, and Amos Tversky.** 1992. "The weighing of evidence and the determinants of confidence" *Cognitive Psychology,* 24: 411-435.

[26] **Grubb, Michael D.** 2009. "Selling to overconfident consumers" *American Economic Review*, 99(5): 1770-1807.

[27] **Henrion, Max, and Baruch Fischhoff.** 2002. "Assessing uncertainty in physical constants." In: *Heuristics and Biases: The Psychology of Intuitive Judgment, ed.* T. Gilovich, Dale Griffin and Daniel Kahneman, 666-677, Cambridge: Cambridge University Press.

[28] **Howell, William C., and Sarah A. Burnett.** 1978. "Uncertainty measurement: a cognitive taxonomy" *Organizational Behavior and Human Performance* 22, 45-68.

[29] **Juslin, Peter,** 1994. "The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items" *Organizational Behavior and Human Decision Processes,* 57: 226-246.

[30] **Juslin, Peter, Andres Winman, and Henrik Olsson.** 2000. "Naive empiricism and dogmatism in confidence research: a critical examination of the hard-easy effect" *Psychological Review*, 107(2): 384-396.

[31] **Kahneman, Daniel and Amos Tversky.** 1982. "Variants of uncertainty" *Cognition*, 11: 143-157.

[32] **Keren, Gideon**. 1987. "Facing uncertainty in the game of Bridge: a calibration study" *Organizational Behavior and Human Decision Processes*, 39: 98-114.

[33] **Klayman, Joshua, Jack B. Soll, Claudia González-Vallejo, and Sema Barlas**. 1999. "Overconfidence: it depends on how, want and whom you ask" *Organizational Behavior and Human Decision Processes*, 79(3): 216-247.

[34] **Koehler, Derek J., Lyle Brenner, and Dale Griffin**. 2002. "The calibration of expert judgment: Heuristics and biases beyond the laboratory. In: *Heuristics and Biases: The Psychology of Intuitive Judgment, ed.* T. Gilovich, D. Griffin and D. Kahneman, 686-715, Cambridge: Cambridge University Press.

[35] **Koellinger, Phillip, Maria Minniti, Christian Schade.** 2007. "I think I can, I think I can: Overconfidence and entrepreneurial behavior" *Journal of Economic Psychology*, 28: 502-527.

[36] **Köszegi, Botond**. 2006. "Ego utility, overconfidence and task choice." *Journal of European Economic Association*, 4(4): 673-707.

[37] **Lewontin Richard C., and D. Cohen**. 1969. "On population growth in a randomly varying environment." *Proceedings of the National Academy of Sciences of USA*, 62: 1056-60.

[38] **Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence D. Phillips**, 1982. "Calibration of probabilities: the state of the art to 1980." In *Judgment Under Uncertainty*: *Heuristics and Biases*, ed. D. Kahneman, P. Slovic, and A. Tversky, Cambridge: Cambridge University Press.

[39] **Louge, Fernando.** 2010. "Contrarianism and Caution in herding models: an evolutionary approach" mimeo, https://sites.google.com/site/fernandolouge/research/herding_old.pdf.

[40] **McNamara, John M**. 1995. "Implicit frequency dependence and kin selection in fluctuating environments" *Evolutionary Ecology,* 9: 185-203.

[41] **Moore**, **Don A.** 2007. "Not so above average after all: when people believe they are worse than average and its implications for theories of bias in social comparison" *Organizational Behavior and Human Decision Processes*, 102: 42-58.

[42] **Moore, Don A., and Paul J. Healy.** 2008. "The trouble with overconfidence" *Psychological Review* 2008, 115 (2): 502-517.

[43] **von Neumann, John, and Oscar Morgenstern.** 1944 (third edition, 1953). *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.

[44] **Odean, Terrance**. 1998. "Volume, volatility, price, and profit when all traders are above average." *The Journal of Finance*, LIII(7): 1887-1934.

[45] **Oskamp, Stuart.** 1965. "Overconfidence in case-study judgments" *Journal of Cognitive Psychology*, 29(3), 261-265.

[46] **Rabin, Matthew, and Joel L. Schrag.** 1999. "First impressions matter: a model of confirmatory bias" *The Quarterly Journal of Economics*, 114: 37-81.

[47] **Rayo Luis, Gary S. Becker.** 2007. "Evolutionary Efficiency and Happiness" *Journal of Political Economy*, 115: 302-336.

[48] **Robson, Arthur J**. 1996. "A biological basis for expected and non-expected utility" *Journal of economic Theory*, 68: 397-424.

[49] **Russo, Edward J., and Paul J.H. Schoemaker**. 1992. "Managing overconfidence" *Sloan Management Rev.* 33: 7-17.

[50] **Samuelson, Paul A.** 1971. "The 'fallacy' of maximizing the geometric mean in long sequences of investing or gambling" *Proc. Nat. Acad. Sci.* 68 , 2493-2496.

[51] **Sandroni, Alvaro, and Francesco Squintani**. 2007. "Overconfidence, insurance and paternalism" *The American Economic Review*, 97(5), 1994-2004.

[52] **Sieber, Joan E**. 1974. "Effects of decision importance on ability to generate warranted subjective uncertainty" *Journal of Personality and Social Psychology*, 30(5): 688-694.

[53] **Skala, Dorota.** 2008. "Overconfidence in psychology and finance - an interdisciplinary literature review" *Bank i Kredy*, 4: 33-50.

[54] **Soll, Jack B.** 1996. "Determinants of Overconfidence and Miscalibration: The Roles of Random Error and Ecological Structure" *Organizational Behavior and Human Decision Processes*, 65(2): 117-137.

[55] **Soll, Jack B., Joshua Klayman**. 2004. "Overconfidence in interval estimates" *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(2), 299-314.

[56] **Svenson, Ola**. 1981. "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychological* 47, 143-148.

[57] **Taylor, Shelley E., Jonathan D. Brown**. 1988. "Illusion and well-being: a social psychological perspective on mental health" *Psychological Bulletin,* 103(2), 193-210.

[58] **Teigen, Karl H., and Magne Jorgensen**. 2005. "When 90% confidence intervals are 50% certain: on the credibility of credible intervals" *Applied cognitive Psychology*, 19: 455-475.

[59] **Van Den Steen, Eric**. 2004. "Rational overoptimism (and other biases)" *American Economic Review*, 94(4): 1141-1151.

[60] **Waldman, Michael.** 1994. "Systematic errors and the theory of natural selection" *American Economic Review*, 84(3): 482-497.

[61] **Wang, F. Albert**. 2001. "Overconfidence, investor sentiment, and evolution" *Journal of Financial Intermediation* 10, 138-170.

[62] **Weinberg, Bruce A.** 2009. "A model of overconfidence" *Pacific Economic Review*, 14(4): 502-515.

[63] **Yates, J. Frank, Ju-Whei Lee, Winston R. Sieck, Incheol Choi, and Paul C. Price.** 2002. "Probability judgment across cultures." In: *Heuristics and Biases: The Psychology of Intuitive Judgment, ed.* T. Gilovich, D. Griffin and D. Kahneman, 271-291, Cambridge University Press.