



MIT Sloan School of Management

**Working Paper 4255-02
June 2002**

SKILL OR LUCK? BIASES OF RATIONAL AGENTS

Eric Van den Steen

© 2002 by Eric Van den Steen. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit including © notice is given to the source.

This paper also can be downloaded without charge from the
Social Science Research Network Electronic Paper Collection:
http://ssrn.com/abstract_id=319972

Skill or Luck?

Biases of Rational Agents

Eric Van den Steen*

June 5, 2002

Abstract

This paper shows why, in a world with differing priors, rational agents tend to attribute their own success more to skill and their failure more to bad luck than an outsider. It further shows why each agent in a group might think he or she is the best, why an agent might overestimate the control he has over the outcome, and why two agents' estimated contributions often add up to more than 100%. Underlying all these phenomena is a simple and robust mechanism that endogenously generates overoptimism about one's own actions. The paper also shows how these biases hinder learning and discusses some implications for organizations.

1 Introduction

People tend to attribute success to their own skills and failures to bad luck (Zuckerman 1979). When asked about the percentage of the household work they perform, a husband's and wife's percentages typically sum up to more than 100% (Ross and Sicoly 1979). More than 80% of US drivers consider themselves better than the median (Svenson 1981). Apart from being fun to know, these observations have real implications for organizations and incentives. It is therefore important for economics to understand their causes and effects. This paper contributes to that effort.

The topic of attribution and inference biases has a long tradition in psychology. Most of that literature (Nisbett and Ross 1980, Fiske and Taylor 1991, Bazerman 1998) uses:

- 'self-serving' bias to refer to the fact that people attribute success to skill and failure to bad luck,

*MIT-Sloan School of Management (evds@mit.edu). John Roberts' suggestions and support made an enormous difference to me. Bengt Holmström's remarks also had a big impact on this and other work. I further got extensive suggestions from Mark Satterthwaite and Jan Zabochnik and very useful comments from Bill Barnett, Max Bazerman, Thomas Hellmann, Ed Lazear, Jon Levin, Sendhil Mullainathan, Paul Oyer, Emily Pronin, Garth Saloner, Antoinette Schoar, Muhamet Yildiz, and the participants in the Stanford Strategy Conference, the Harvard-MIT Theory Seminar, the MIT Organizational Economics Lunch, and the SITE conference. Financial support from the Interuniversitair Centrum voor Managementwetenschappen (ICM/CIM) and Stanford Graduate School of Business are gratefully acknowledged.

- ‘ego-centric’ or ‘self-centric’ bias to refer to the fact that people in a joint enterprise relatively over-estimate their contribution (to a good outcome),
- ‘overconfidence’ to refer to the fact that people over-estimate the precision of their estimates and predictions, and
- ‘illusion of control’ to refer to the fact that people overestimate the influence they have over the outcome of a partially random event.

To avoid confusion with this existing terminology, I will furthermore use

- ‘overoptimism’ to refer to the fact that people tend to be overoptimistic about future life events and the success of their actions, and
- ‘self-enhancing’ bias to refer to the fact that people consider themselves better than they really are.

Most of the psychology literature¹ has interpreted these effects as motivational biases (Zuckerman 1979, Baumeister 1998): people hold unrealistically positive views of themselves since doing so increases their happiness and general well-being. This is essentially a theory of wishful thinking. Some authors in the psychology literature have challenged that view by forwarding cognitive explanations, in particular for the self-serving bias. Miller and Ross (1975), for example, cite experimental evidence that people are more likely to accept responsibility for expected outcomes and that people generally expect to succeed. They note that these two combine to a self-serving bias.

This paper shows that even perfectly Bayesian-rational agents will tend to suffer from these biases if we allow for differing or heterogeneous priors. The mechanisms are quite different from these suggested in the psychology literature and therefore have also different implications and remedies.

Central to the theory is a mechanism that leads endogenously to an agent being relatively overoptimistic about the likelihood of success of her own actions. To see how this comes about, consider a situation in which agents differ in their beliefs about what actions are most likely to succeed. Each agent selects her actions based on her own beliefs, but also uses these same beliefs to evaluate her own and others’ actions. This double use of the same set of criteria favors the focal agent’s actions in the evaluation stage. In particular, the agent will tend to conclude that her own actions are more likely to succeed than these of others. Note, however, that this also works the other way: the other agents will tend to consider the focal agent to be overoptimistic about her own actions. A different way of seeing this is the following. From an outsider’s perspective (or from the reference perspective) the focal agent is mistaken in her assessment of the probabilities of success of different actions. When trying to select the action with the highest probability of success, the focal agent is typically more likely to select actions of which she relatively overestimated the probability of success than those of which she underestimated that probability. Overall, the agent will thus tend to be

¹A more complete literature overview, including economic contributions, follows below.

overoptimistic from that outsider’s perspective. Random variations plus systematic choice lead to systematic variations.

Such overoptimism causes the ‘skill versus luck’ attribution mentioned above. To see how, consider a Bayesian agent who believes that the action he has chosen is nearly certain to succeed. In the event of a failure, he will attribute the failure to a stroke of bad luck rather than to a bad choice of action since he was so sure the action was correct. An outside observer, on the other hand, who had less optimistic beliefs about the action, will tend to put the blame relatively more on the choice of action. This is exactly the self-serving bias. Probably the most interesting conclusion is that such bias also leads to insufficient learning, since the decision maker will discount his failures as random variations and consider them irrelevant. Own actions thus constitute a biased experiment.

The overoptimism also naturally causes each agent to consider himself to be better than the median. Note how this would translate to an academic context: each researcher would consider his own projects to be more relevant than these of his colleagues, but only because he thought they were the most relevant to start with and is therefore working on them. Finally, in cooperative ventures, the focal agent will tend to consider his own contribution to be more important than that of his colleagues.

Note that the paper does not shift the problem by assuming the biases directly in the priors: it derives a systematic bias from random but unbiased variations. The theory also has empirical content. As will be shown later, for example, the biases increase as the number of alternative actions increases. This prediction is different from that of the alternative ‘wishful thinking’ theory in psychology (or any other theory, to my knowledge), in which there is no reason for the bias to depend on the number of non-chosen alternatives.

Note also that agents in this model are not assumed to somehow know the ‘true’ prior. Instead, they (knowingly) entertain differing beliefs, i.e. they start from differing or heterogeneous priors. An important implication is that agents can openly disagree and do *not* necessarily update their beliefs just because they see that others disagree with them. They just think these others are wrong². Section 2 gives a concise motivation and discussion of this approach. While many of the results are then stated most naturally in subjective terms, I will sometimes designate one set of beliefs as the ‘reference’ beliefs. These can be interpreted as the beliefs of the researcher or of the social planner, or even as the ‘objective’ but unknown probabilities (although this latter interpretation may cause confusion).

Overall, an interesting way to state the key result of this paper is the following: as soon as we relax the strong assumption that people somehow have identical and true priors, many of the well-known behavioral biases appear very naturally. It is not the purpose of this paper,

²The idea that people who observe that others disagree with them conclude that these others must be wrong is in line with Harsanyi’s observation that ‘by the very nature of subjective probabilities, even if two individuals have exactly the same information and are at exactly the same high level of intelligence, they may very well assign different subjective probabilities to the very same events’. It is of course possible to enrich the current model with private information, in which case agents will update their beliefs (partially) when they see that others disagree with them. As long as the priors differ, however, (almost surely) no amount of updating will eliminate the disagreement. The effects discussed in this paper will thus persist. For simplicity, the analysis in this paper therefore assumes common knowledge of the disagreement, i.e. it assumes that all such updating has already taken place, and abstracts from private information.

however, to suggest that all apparent biases can be explained in Bayesian rational terms. As I will discuss at various points, some of the experimental results cannot be explained by the mechanisms discussed here. This does, however, *not* deny the value of studying ‘rational’ mechanisms that may be playing. The study of such mechanisms might allow us to predict at least part of the bias, how it is affected by the structure of the task, what its implications are for organizations, and to assess potential remedies.

The endogenous overoptimism effect in this paper is structurally very similar to the winner’s curse (Capen, Clapp, and Campbell 1971), regression towards the mean (Kahneman and Tversky 1973), post-decision surprises (Brown 1974, Harrison and March 1984), and Lazear’s (2001) explanation for the Peter Principle (Peter and Hull 1969). Brown (1974), for example, discusses how this kind of mechanism causes investment projects to fall short of their estimated revenue. Within the sociology literature, Harrison and March (1984) make an analogous argument with a semi-formal model. There is of course a broad economic literature on the winner’s curse under common priors, starting with Wilson (1977). Compte (2002) shows that the winner’s curse also applies to the independent private values case. Finally, Van den Steen (2002) uses a mechanism similar to the one in this paper to show how rational agents might become overconfident.

The self-serving bias has been studied in economics for its role in bargaining and litigation (Priest and Klein 1984, Babcock, Loewenstein, and Issacharoff 1995, Babcock and Loewenstein 1997). Research on the potential causes and its implications in other areas is limited. Zabojnik (2000) presents a model in which people can learn about their abilities at a cost and concludes that in the limit a proportion of the population ends up being overoptimistic about their abilities, i.e. displaying a self-enhancing bias. There will, however, also be a proportion that is under-confident, since the beliefs follow a stopped martingale. In the current paper, in contrast, all agents almost surely end up with a self-enhancing bias.

There is further a small but growing literature that shows how overoptimism or overconfidence can be generated by other forms of bounded rationality, such as selective recall. Benabou and Tirole (1999), for example, show how self-serving biases can arise if people have (endogenously chosen) imperfect recall and the recalled information is always bad news. Rabin and Schrag (1999) show how a confirmatory bias may lead to overconfidence. Compte and Postlewaite (2001) show that self-serving biases in interpreting or remembering outcomes may be optimal if an agent’s performance depends on his self-confidence.

With respect to the economic literature, the key contribution of this paper is to show that many behavioral biases appear quite naturally as soon as we allow agents to have differing or heterogeneous priors, and to formally link overoptimism to self-serving, egocentric, and self-enhancing biases, and to an illusion of control. Except for explaining these biases, the theory also generates some new conclusions such as how the self-serving bias hinders learning, why group decisions make less of an improvement than might seem, and why experimental results often underestimate the true impact of behavioral biases. A short discussion suggests many organizational implications.

With respect to the sociological and psychological overoptimism-related literature, the contribution of this paper is to bring the phenomenon in the domain of formal economic analysis by providing a theory built on a Bayesian foundation, in the sense of differing priors.

This permits a very general approach, with comparative statics and robustness checks, and provides a solid foundation for studying further implications. One immediate result is the link among the different biases. As a different application I will also consider joint decision making.

Sections 2 and 3 will provide some groundwork for the main part of the paper. The first motivates and discusses the use of heterogeneous or differing priors, while the second discusses some of the existing experimental evidence on self-serving and related biases. Section 4 starts the analysis by formalizing the endogenous overoptimism effect. It also considers comparative statics and robustness. Section 5 derives the self-serving bias and studies its implications for learning. That section ends with an extension of the original model in which rational people also over-estimate the control they have over the outcome. Section 6 considers how these models translate in cooperative projects with each player taking the praise for himself and putting the blame on the other. Section 7 studies a variation on the main model of the paper in which each person considers himself the best driver of the population. Section 8 finally discusses organizational implications. Some of the proofs are in appendix C.

2 A note on ‘differing beliefs’ in economic modeling

The models in this paper differ in one respect from most of the economic literature: the agents knowingly entertain differing beliefs³ (without having private information). The reason for this assumption is pragmatic: differences in beliefs are at the heart of the issues studied here, and assuming common knowledge of differing beliefs is the most transparent and parsimonious way to study this question. Differing beliefs do not contradict the economic paradigm: while rational agents should use Bayes’ rule to update their prior with new information, nothing is said about those priors themselves, which are primitives of the model. In particular, absent any relevant information agents have no rational basis to agree on a prior. Harsanyi (1967), for example, observed that ‘by the very nature of subjective probabilities, even if two individuals have exactly the same information and are at exactly the same high level of intelligence, they may very well assign different subjective probabilities to the very same events’. The best argument for the alternative assumption that agents share a common and correct prior is Aumann’s (1987) argument that making this assumption allows us to ‘zero in on purely informational issues’. Conversely, differing priors allow us to zero in on the implications of open disagreement and differing beliefs.

Van den Steen (2001) considers this issue in more detail⁴. Among other things, it argues against the idea that differing priors might allow us to explain anything, it discounts the theoretical possibility that agents will make very large or infinite bets, and shows that the epistemic foundations for Nash equilibria in the sense of Aumann and Brandenburger (1995)

³Whenever we refer to belief differences, or agents entertaining differing beliefs, we mean that ‘agents have differing beliefs about a specific event and their beliefs are common knowledge’. In economics, the term ‘disagreement’ is often used to denote such belief differences. We avoid this term since it suggests conflict. This paper is definitely not the first use of such differing priors. See for example Harrison and Kreps (1978) or Yildiz (2000).

⁴See also Morris (1995) or the discussion between Gul (1998) and Aumann (1998).

extend to this context with differing priors on the payoff-space.

Section 4 discusses concisely in how far the results of this paper can be obtained in a common priors context.

3 Experimental Evidence on Self-serving and Related Biases

Most of the (extensive) experimental research on the self-serving bias was done between the mid-sixties and the early eighties. Two surveys, Miller and Ross (1975) and Zuckerman (1979), are still widely cited for setting the perimeter of the essential experimental conclusions. More recent research has confirmed these earlier results and tested specific hypotheses on the underlying causes and interactive variables. To get a feel for this type of experimental research, consider the following study by Streufert and Streufert (1969). Pairs of subjects played an international simulation game against other teams. Manipulated feedback suggested that the pair either improved or became worse over time. The subjects were then asked to attribute their performance either to the own team or to the competing team. It turned out that subjects attributed more responsibility to their own team when they improved than when they did not.

While this particular study had a fairly clean result, the overall literature is not so coherent or transparent. Differences in experimental setups, procedures, and potentially confounding factors make it difficult to compare the findings. Nevertheless, a relative consensus has emerged (Fiske and Taylor 1991, Baumeister 1998, Duval and Silvia 2002) which holds that:

- There is clear and fairly consistent evidence of a positive self-serving bias (people claim credit for success). There is also evidence for a negative or self-protective bias (people attribute failure to exogenous causes), though the latter seems to be more sensitive to context.
- Both motivation (the desire to see oneself in a flattering light) and cognitive limitations drive the self-serving bias, but the motivational causes are more prevalent.

Large parts of the literature also agree that:

- The self-serving bias is robust cross-culturally (Fiske and Taylor 1991).
- Inconsistencies in the evidence are partially driven by conflicts between motivation (the desire to see oneself in a positive light) and cognition (what the agents know) (Pyszczynski and Greenberg 1987).
- People with high self-esteem show more self-serving biases and tend to exaggerate their control over events, especially over those that turn out to be successful. People with low self-esteem sometimes even have a self-deprecating bias (Blaine and Crocker 1993).

- The self-serving bias is systematically influenced by mediating factors such as gender of the subjects or the prospect of being evaluated at a later stage (Zuckerman 1979).

The end of section 5 considers how well the current model fits this evidence. It is useful, however, to give already two caveats. First of all, the introduction mentioned already that the mechanism in this paper is probably not the only one. Conclusions that there are also, for example, motivational causes do therefore not necessarily contradict the theory. Second, most of the experimental studies in this area have people solve problems like anagrams. This is quite different from the task of choosing a course of action, which this paper focuses on. There is, however, also clear evidence of the self-serving bias in contexts that are similar to the current one, such as Streufert and Streufert (1969) or Curren, Folkes, and Steckel (1992).

Other Biases The evidence for the self-enhancing bias (people have inflated views of themselves) is extensive and covers many categories and situations (Taylor and Brown 1988). There is also evidence that this bias increases as the criteria are less specific (Allison, Messick, and Goethals 1989, Wade-Benzoni, Thompson, and Bazerman 2000).

The evidence for overoptimism is also quite substantial (Weinstein 1980, 1982, 1984 and the reviews therein). Weinstein (1984), for example, asked college students to rate their chances, relative to those of their peers, to experience specific health problems and analyzed the reasons they cited for their assessment. When the subjects considered their own actions to be the driving force of their risk rate, they tended to be overoptimistic, but not so when they considered their hereditary background or environmental conditions to be the driving force. An earlier study (Weinstein 1982) came to an analogous conclusion but also identified other contributing factors, such as prior experience with the event.

The evidence on egocentric biases is more limited. The above cited study by Ross and Sicoly (1979) on married couples is the most well-known. Wolosin, Sherman, and Till (1973) also found an egocentric bias in a coordination experiment, while Sedikides, Campbell, and Reeder (1998) found the bias in distant relationships but not in close relationships.

Langer (1975) showed that people valued lottery tickets higher when they had personally chosen them than when they were randomly assigned to them. She termed this the ‘illusion of control’. While the effect has been confirmed on other occasions, the literature on the topic is also fairly limited to my knowledge.

4 Endogenous Overoptimism

Let agents sometimes overestimate and sometimes underestimate the probability of success of an action. When trying to select the action with the highest probability of success, they are more likely to select actions of which they overestimated the probability of success than those of which they underestimated that probability. Overall, agents will thus tend to be overoptimistic about the likelihood of success of the actions they undertake. Since this mechanism underlies much of what is going on in this paper, we start the formal analysis by studying it in detail. This section analyzes first the basic overoptimism mechanism and its

comparative statics and then considers some important issues, such as robustness and the role of differing priors.

4.1 Basic analysis

Consider I agents who each have to decide what action to undertake. Let there be N potential actions A_n with $1 \leq n \leq N$. Each agent's outcome can be either a success or a failure, denoted S and F respectively. In choosing the action, the agents try to maximize the likelihood of having a success⁵.

An action's probability of success is uncertain. Agents have subjective beliefs about this and these beliefs are commonly known⁶. In particular, let p_n^i denote agent i 's belief regarding the probability of success of action n . Assume that the p_n^i are i.i.d. draws from a distribution F on $[\underline{p}, \bar{p}] \subset [0, 1]$, i.e. $p_n^i \sim F[\underline{p}, \bar{p}]$ with $\underline{p} < \bar{p}$. Let F be atomless. Note that F is *not* a prior and agents do *not* have any private information. We analytically use F to represent the empirical variation in beliefs of agents, but F is not known to those agents, but only to us, the analysts. Even if an agent knew F , that wouldn't tell him anything since F does not contain information.

This set-up can also be interpreted as follows. An action is either right or wrong. Right actions always succeed, while wrong actions always fail. There is uncertainty whether a particular action is right or wrong and p_n^i denotes agent i 's belief that action n is right.

Most of the analysis that follows is specified completely in subjective terms, in the sense that it says how one agent considers the judgment of another. Alternatively, we can introduce a 'reference' probability of success for action n , denoted p_n . This can be interpreted as the belief of the researcher (in case of an experiment) or of the social planner (for normative purposes) or even as the objective truth (though the latter interpretation is sometimes confusing). To obtain meaningful results, we should impose some relationship between this reference belief and the beliefs of the agents. A purely subjective approach treats the reference belief as that of any other agent. This implies that we should assume

Assumption 1 *The reference belief p_n is distributed according to F , independent from p_n^i .*

Appendix B shows that the key results still obtain when the agents' beliefs depend on the reference belief.

The first result of this paper is that agents are relatively overoptimistic about their likelihood of success. Let Y_i denote the project chosen by agent i . Let $p_{Y_i}^j$ denote the belief of some agent j that Y_i will be a success (where we allow $j = i$). Let finally $\rho_i^j = p_{Y_i}^j - p_{Y_i}^i$ and $\rho_i = p_{Y_i}^i - p_{Y_i}$ denote i 's overoptimism about her actions relative to, respectively, agent j and the reference probability.

Proposition 1

⁵In particular, one can assume that agents are risk-neutral revenue maximizers and that a success gives a payoff 1, while failure gives 0 payoff.

⁶Common knowledge is not necessary for most of the results, but it makes explicit that the biases will not disappear simply because agents are made aware of them.

- When the number of actions $N > 1$, $\forall j \neq i$, $E[\rho_i^j] > 0$: In expectation, every agent other than i considers i to be strictly overoptimistic about her own project (in the sense of over-estimating the probability of success).
- $\forall j \neq i$, $\rho_i^j > 0$ a.s. as $N \rightarrow \infty$: All agents other than i almost surely consider i to be overoptimistic in her own project, when the number of alternative projects $N \rightarrow \infty$.
- If the maximum of the support $\bar{p} = 1$ then $\forall \epsilon_1, \epsilon_2 > 0$, $\exists \tilde{N}$ such that $\forall N \geq \tilde{N}$, $P[p_{Y_i}^i > 1 - \epsilon_1] > 1 - \epsilon_2$: With $N \rightarrow \infty$, any particular agent i will almost surely believe that her preferred project Y_i will almost surely succeed.
- Under A1, $E[\rho_i] > 0$ when $N > 1$ and $\rho_i > 0$ a.s. as $N \rightarrow \infty$. In expectation (and a.s. when $N \rightarrow \infty$) the agent is overoptimistic (in a reference sense) about the likelihood of success of her project.

Proof : Without loss of generality, let $i = 1$. Note that, by integration by parts, $E[p_{Y_i}^i - p_{Y_i}^j] = \int_{\bar{p}}^{\bar{p}} F(x) - F^N(x) dx$. This equals zero when $N = 1$ and is strictly larger than zero when $N > 1$, which proves the first part. For the second part of the proposition, the probability that all agents consider agent 1 to be overconfident is $P[p_{Y_1}^1 \geq p_{Y_1}^2, \dots, p_{Y_1}^1 \geq p_{Y_1}^N] = \int P[x \geq p_{Y_1}^1]^{I-1} f_{p_{Y_1}^1}(x) dx = \int F(x)^{I-1} N F(x)^{N-1} dF(x) = \frac{N}{N+(I-1)}$ which converges to 1 as $N \rightarrow \infty$. For the third part, note that with $\bar{p} = 1$, $P[p_{Y_1}^1 > 1 - \epsilon_1] = 1 - F_{p_{Y_1}^1}(1 - \epsilon_1) = 1 - [F(1 - \epsilon_1)]^N$. Since $\epsilon_1 > 0$, $1 - \epsilon_1 < \bar{p}$ so that $F(1 - \epsilon_1) < 1$. This implies that $\lim_{N \rightarrow \infty} 1 - [F(\bar{p} - \epsilon_1)]^N = 1$.

The last part is a direct consequence of the others. ■

Note that, following the proof of the proposition, $E[\rho_i^j] = 0$ when the number of actions $N = 1$: The agent will thus only be overoptimistic when she has a choice of action. This highlights the key role of choice in this model. This remark has at least three important implications. First of all, whenever agents are overoptimistic about their projects but did not have a choice of action, a different mechanism than the one described here must be at work. Second, by eliminating choice, it is possible to eliminate the overoptimism generated by this mechanism. This suggests, for example, the use of outsiders to estimate the likelihood of success after a project has been selected. Third, experimental studies that limit the agents' freedom of action, as they often do, will underestimate the magnitude of the effect.

Note also the overall situation :

- Agents have a lot of confidence in their own projects and less so in projects of others.
- Each agent thinks all other agents have too little confidence in her own project and too much confidence in their own projects.
- Subjective rationality can lead to objective overoptimism.

It is very tempting to translate these results to an academic context. In particular, let the actions correspond to research topics. Let each academic be limited to choosing one research topic. Let the contribution to science of any project be measurable on a scale from zero to one. In particular, assume that a successful project contributes one unit to science while a failure contributes zero. For simplicity, make the (heroic) simplifying assumption

that the probability of success of a research project depends only on the project type and not on the researcher.

Corollary 1 *In the limit as the number of potential research topics goes to infinity, each academic considers his own expected contribution to be among the highest (i.e. to converge to 1) while he considers the contribution of others to be distributed according to F on $[0, 1]$. In particular, he considers his own contribution to be above the median.*

The inflated view of own work that academics sometimes observe among their colleagues (Ellison 2000) might thus be a matter of genuinely differing priors.

Comparative Statics Although the model is very simple, it allows some interesting comparative statics. The first of these shows that overoptimism increases in the number of available projects. It makes again clear how the freedom to choose actions is the driving force behind the overoptimism, with more freedom causing more overoptimism.

Proposition 2a

- $\frac{dE[\rho_i^j]}{dN} > 0$ and under A1 $\frac{dE[\rho_i]}{dN} > 0$: The expected level of overoptimism strictly increases in the number of available projects N .

Proof : By an earlier argument, $E[\rho_i^j] = E[p_{Y_i}^j - p_{Y_i}^j] = \int_{\underline{p}}^{\bar{p}} F(x) - F^N(x)dx$. Since $\forall x \in (\underline{p}, \bar{p})$, $F^N(x)$ decreases strictly in N , the first part immediately follows. The second part is analogous. ■

As mentioned earlier, this result bears on the conclusions from experiments. Since laboratory setting typically restrict the degrees of freedom of the agent, they might underestimate the importance of overoptimism.

The other comparative statics relate to the distribution function. The main result here is that overoptimism increases in the underlying uncertainty, but we actually obtain results that relate to moments of any order (including skewness and kurtosis). Remember that G is a mean-preserving spread of F if, for F and G having support contained in $[0, 1]$, $\int_0^1 G(y) - F(y)dy = 0$ while $\int_0^x G(y) - F(y)dy \geq 0 \forall x$ with $0 < x < 1$ with strict inequality for some x . In that case, $\mu_F = \mu_G$ while $\sigma_G^2 > \sigma_F^2$. This is a special case of ‘lower moment preserving stochastic dominance’, defined as follows. Let $F_0(x) = F(x)$ and $F_k(x) = \int_0^x F_{k-1}(t)dt$. Say now that F ‘lower moment preserving m -order stochastically dominates’ G , denoted $F \text{ MPSPD}_m G$, if

1. $F_k(1) = G_k(1)$ for $k = 1, \dots, m - 1$
2. $F_{m-1}(x) \leq G_{m-1}(x)$ with strict inequality for at least one x

Appendix A shows that in this case, all moments up to m are identical while the m ’th moment of F is larger than that of G if m is odd, and smaller if it is even.

Proposition 2b

- If G is a mean-preserving spread of F , then $E_G[\rho_i^j] > E_F[\rho_i^j] \forall N > 1$ and $\rho_{i,G}^j > \rho_{i,F}^j$ a.s. when $N \rightarrow \infty$: The expected level of overoptimism increases in the underlying uncertainty.

- If F and G are analytic functions and $F \text{ MPD}_m G$ for some even $m \geq 2$, then $\exists \tilde{N} \text{ s.t. } \forall N \geq \tilde{N}, E_G[\rho_i^j] > E_F[\rho_i^j]$. Moreover, as $N \rightarrow \infty$ $\rho_{i,G}^j \geq \rho_{i,F}^j$ a.s.. When m is odd, the opposite inequalities hold.

The last comparative static follows essentially from the fact that the upper tail distribution is what determines overoptimism.

4.2 Further considerations

Comparison with the experimental evidence The mechanism discussed in this section fits the earlier mentioned experimental fact that people tend to be overoptimistic about future life events. In particular, the comparative static of proposition 2a could explain why people are especially overoptimistic when they feel that the event is influenced by their actions (Weinstein 1984). The fact that the overoptimism tends to increase as the person has less experience with the event (Weinstein 1982), is consistent with the comparative static of proposition 2b if we assume that less experience causes a mean-preserving spread. But not all the evidence fits easily. The fact that people are still overoptimistic even when they have no choice, for example, has currently no place in the model.

The need for differing priors A natural question is whether we need to allow for differing priors to explain these outcomes. In particular, it may seem that even with common priors an agent will tend to choose the actions of which he overestimated the probability of success. In that case, however, the agent will be aware of his tendency, and take it into account in his inferences and decisions. In particular, in such models it cannot happen that on average more than 50% of the agents consider themselves to be better than the median. Most empirical studies on these topics, however, obtain exactly that type of results. They typically ask the subjects to estimate in which percentile they fall and conclude that agents are overoptimistic in this strong sense. Such outcomes only fit a model with differing priors. This remark applies in nearly identical form to the subsequent sections and will therefore not be repeated there.

Note also that adding private information to the model would make it look more realistic, but would not change the basic results. In particular, upon observing that there is disagreement, the agents would partially update their beliefs. With differing priors, however, the disagreement will (almost surely) never disappear and the effects discussed in this paper will thus persist. For the benefit of transparency, I therefore abstract here from any private information.

Robustness The results up to this point assume that the beliefs are independent draws from a common distribution. This poses the issue how robust they are to alternative specifications.

To get a perspective on this robustness issue, consider a situation with 2 agents and focus on agent 1. Let p_n^1 and p_n^2 denote the beliefs of the respective agents regarding action n . Let F denote the *joint* distribution function of their beliefs and let $S \subset \mathbb{R}^2$ be the

support of F . Assume that S has a non-empty interior. Define $\bar{p}^1 = \max_{(p_n^1, p_n^2) \in S} p_n^1$ and $\bar{p}^2 = \max_{(p_n^1, p_n^2) \in S} p_n^2$.

Proposition 3 *As the number of actions $N \rightarrow \infty$, agent 1 is almost surely overoptimistic ($p_{Y_1}^1 > p_{Y_1}^2$ a.s.) under any of the following conditions:*

- $\bar{p}^1 > \bar{p}^2$, or
- p_n^1 and p_n^2 are distributed independently with densities f_1 and f_2 and $\bar{p}^1 \geq \bar{p}^2$, or
- p_n^1 and p_n^2 have joint distribution F with continuous density $f > 0$ on S and $\bar{p}^1 \geq \bar{p}^2$.

While this suggests that the basic result is quite robust, it is not completely so. For a counterexample, consider the following situation. For some action n , let the true probability of success p_n be drawn from a uniform distribution on $[0, 1]$. Each agent observes $p_n^i = p_n + \epsilon_i$ with ϵ_i a mean-zero distributed random variable which may depend on the value of p_n . In this situation, the agent will in expectation still be overconfident for any number of alternative projects $0 < N < \infty$. The bias, however, converges to zero as $N \rightarrow \infty$. What happens in this case is that as $p_n \rightarrow 1$, for $p_n + \epsilon_i$ to be unbiased, it must be that the density of ϵ_i on $[p_n, 1]$ goes to infinity. So upon selecting an action with a very high subjective probability of success p_n^i , the underlying value of p_n goes almost surely to one. Note that this case does not satisfy the conditions of assumption 2.

A general utility-overestimation theorem The logic behind this theory holds in fact in more generality than described above. Consider a general choice problem with I symmetric agents. Each agent's ex-post utility $u(x, y)$ depends on his choices x and the state of the world y . The state of the world is described by $y \in Y$, with Y some general state-space. The state y is unknown to the agents, but each agent i has his subjective belief, described by a probability measure μ_i . Let \mathcal{M} denote the space of all such probability measures. The beliefs of any two agents i and j are jointly distributed according to a measure $F_{i,j}$ over $\mathcal{M} \times \mathcal{M}$. Assume that $F_{i,j}$ is symmetric in the sense that $F_{i,j}(\mu, \nu) = F_{i,j}(\nu, \mu)$. Agent i 's choices are an element $x \in X$ with X some general choice space. Altogether, this implies that agent i 's expected utility according to himself is

$$E_i u_i(x) = \int u(x, y) d\mu_i(y)$$

Note that the utility does not depend on the agent, except through his belief μ_i and his choice of action x . Denote by \hat{x}_{μ_i} i 's optimal choice, given these beliefs. Assume that for any belief, such optimal action exists. This gives then the following general theorem⁷

Proposition 4 $E_{F_{i,j}}[E_i u_i(\hat{x}_{\mu_i}) - E_j u_i(\hat{x}_{\mu_i})] \geq 0$: *On average, agent i overestimates his expected utility according to agent j .*

The inequality will be strict when there exists a set $M \subset \mathcal{M}$ with $M \times M$ having strictly positive measure under $F_{i,j}$ and such that for each belief $\mu \in M$ there is a unique optimal action and that action differs from the optimal action of any other belief $\nu \in M$, $\nu \neq \mu$.

⁷This theorem was suggested by Bengt Holmström.

Proof : Let first $\mu_i = \mu$ and $\mu_j = \nu$. Since utility depends on the agent only through his beliefs, the optimal actions under μ and ν are identical for all agents. Denote the respective optimal actions by \hat{x}_μ and \hat{x}_ν . $E_i u_i(\hat{x}_{\mu_i}) = \int u(\hat{x}_{\mu_i}, y) d\mu_i(y) = \int u(\hat{x}_\mu, y) d\mu(y)$ and $E_j u_i(\hat{x}_{\mu_i}) = \int u(\hat{x}_{\mu_i}, y) d\mu_j(y) = \int u(\hat{x}_\mu, y) d\nu(y) \leq \int u(\hat{x}_\nu, y) d\nu(y)$ so that $E_{F_{i,j}}[E_i u_i(\hat{x}_{\mu_i}) - E_j u_i(\hat{x}_{\mu_i})] \geq \int [\int u(\hat{x}_\mu, y) d\mu(y)] - [\int u(\hat{x}_\nu, y) d\nu(y)] dF_{i,j}(\mu, \nu)$. With now $\mu_i = \nu$ and $\mu_j = \mu$ and using the symmetry of $F_{i,j}$, we get $E_{F_{i,j}}[E_i u_i(\hat{x}_{\mu_i}) - E_j u_i(\hat{x}_{\mu_i})] \geq \int [\int u(\hat{x}_\nu, y) d\nu(y)] - [\int u(\hat{x}_\mu, y) d\mu(y)] dF_{i,j}(\mu, \nu)$. Summing both inequalities gives indeed $E_{F_{i,j}}[E_i u_i(\hat{x}_{\mu_i}) - E_j u_i(\hat{x}_{\mu_i})] \geq 0$. That we get a strict inequality follows since, under the specified conditions, $\int u(\hat{x}_\mu, y) d\nu(y) < \int u(\hat{x}_\nu, y) d\nu(y)$. ■

Note that this theorem is essentially an implication of revealed preference.

Joint decision making An obvious question is whether the bias can be eliminated by involving multiple people in the decision process. The answer is actually not so obvious. In particular, consultants often claim that client involvement in the process is effective at increasing the client's buy-in, i.e. their personal belief that the decision is correct and thus their willingness to implement it, but not really at improving the quality of the decision itself. The current model suggests how that might actually make some sense. The intuition is simply that the agents will tend to select a project on which they are all overconfident. The analysis below shows in particular that as the number of alternative projects becomes very large, the agents become as overoptimistic about their joint choice as a single decision maker would be about his own choice.

To see this formally, consider again the basic model of this section with I agents. Assume that 2 of those agents, say 1 and 2, will jointly choose one project. To keep the analysis as general as possible assume that this choice is made as follows. Let f be a function with domain $[0, 1] \times [0, 1]$ and range $[0, 1]$ that increases strictly in both its arguments. Assume that the agents choose the project with the largest $p_n^x = f(p_n^1, p_n^2)$. The key is now that in the limit as there are many projects, the agents are, from the perspective of an outsider, individually and on average equally overoptimistic as an agent who chooses a project on his own.

To state this more formally, let Y_x denote the project that the agents 1 and 2 choose when they choose jointly. Let Y_3 be the project that some agent 3 chooses. Let agent 0 be the outsider. Let, as before, ρ denote the level of overoptimism, so that $\rho_3^0 = p_{Y_3}^3 - p_{Y_3}^0$ and $\rho_1^0 = p_{Y_x}^1 - p_{Y_x}^0$ and $\bar{\rho}^0 = \frac{p_{Y_x}^1 + p_{Y_x}^2}{2} - p_{Y_x}^0$.

Proposition 5 As $N \rightarrow \infty$, $E[\bar{\rho}^0] \rightarrow E[\rho_3^0]$ and $E[\rho_1^0] \rightarrow E[\rho_3^0]$

Proof : Let the maximum of the support of p_n^x be \bar{p} and let μ_F denote the mean of F . As $N \rightarrow \infty$, $\bar{p}_{Y_x} \xrightarrow{\text{a.s.}} \bar{p}$. This implies that $p_{Y_x}^1 \xrightarrow{\text{a.s.}} \bar{p}$ and $p_{Y_x}^2 \xrightarrow{\text{a.s.}} \bar{p}$, so that $E[\rho_1^0] = E[p_{Y_x}^1 - p_{Y_x}^0] \rightarrow \bar{p} - \mu_F$. The same holds for all other ρ , which implies the proposition. ■

Note that this setup does not allow agents to integrate their priors. Whether and how group processes might actually lead to changes in priors is an interesting question, but one that goes beyond the Bayesian axiomatization used in this paper.

5 Skill or luck?

As mentioned earlier, people tend to attribute success to their own skills but to blame failure on bad luck (Zuckerman 1979). This section builds on the overoptimism result of section 4 to show how perfectly rational agents may come to exhibit such self-serving bias. Subsection 5.2 shows furthermore how rational agents might also overestimate the degree of control they have over the outcome.

First, however, we need to state the phenomenon a bit more precisely. In particular, nearly any outcome of interest is simultaneously influenced by both the choice of action and random factors. The above attributions to skill or luck, instead of being categorical, thus refer to the likelihood or degree to which each factor contributed to the outcome. The model operationalizes this as follows. With some probability, success is completely random and independent of the action chosen. This case is denoted ‘pure luck’ or simply ‘luck’. With the complementary probability, the choice of action is relevant and determines the probability of success. This case is denoted ‘skill’ since the likelihood of success now depends on a skillful choice of action. The ex-post attribution is about the likelihood that the outcome was determined by pure luck or influenced by the choice of action. The section on robustness considers a much more general but less transparent implementation.

Consider now the basic setup : there is some chance that the outcome will unobservably be determined by pure luck rather than probabilistically by the agent’s action⁸. The inferred ex-post probability that it was indeed pure luck which caused the success, follows from Bayes’ law. Let, in particular, ‘Action’ denote the event that the agent’s action influenced the outcome and let q indicate the prior probability of ‘Action’. Let further $P_{S|A} = P[\text{Success} | \text{Action}]$, $P_{S|L} = P[\text{Success} | \text{Pure Luck}]$, and $P_{A|S} = P[\text{Action} | \text{Success}]$. Bayes’ law implies

$$P_{A|S} = \frac{P_{S|A}q}{P_{S|A}q + P_{S|L}(1 - q)} \quad (1)$$

This shows that, as the probability of success of the agent’s action $P_{S|A}$ increases, it is ex-post more likely that a success was due to a good choice of action ($P_{A|S}$). Analogously, the ex-post probability that failure was caused by pure (bad) luck also increases. It follows that a rational agent who assesses ex ante $P_{S|A}$ to be higher will tend to attribute success more to skill and failure more to bad luck. The asymmetric attribution of success and failure is thus *in itself* not biased *if* the agent believes his action has a higher probability of success than ‘pure luck’. It follows that studies such as the one by Streufert and Streufert (1969) mentioned earlier, do not contradict in any sense rationality or common priors. The current analysis looks therefore for the more sophisticated version of the self-serving bias: the fact that the focal agent attributes success more to skill (and failure more to bad luck) than an outside observer.

Section 4 showed, in particular, how rational agents tend to be overoptimistic about the actions they undertake. This combines with the Bayesian inference above to produce indeed the overall effect that rational people will tend to attribute their own success more to skill than outsiders do and their own failures more to bad luck than outsiders do. This

⁸A more precise formulation of the game follows.

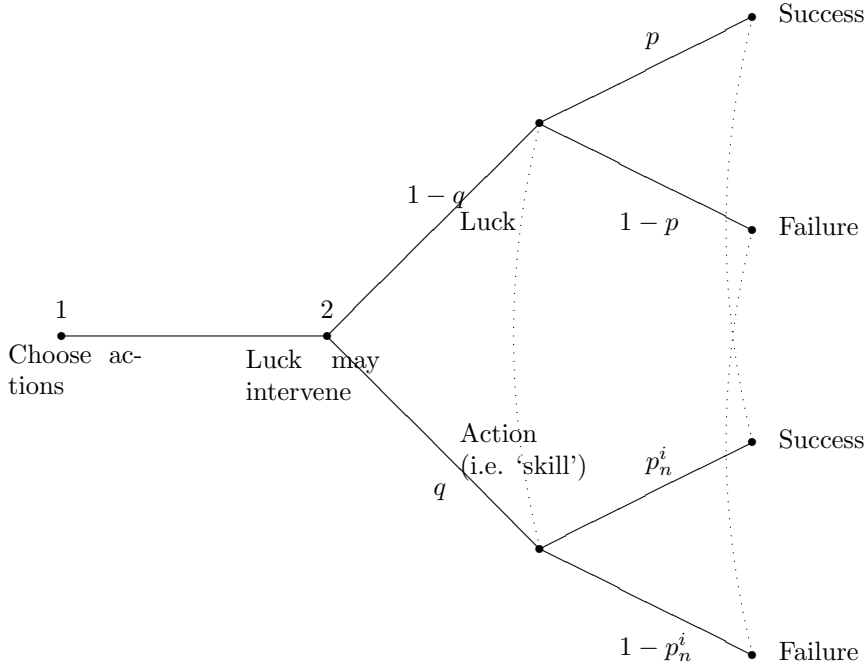


Figure 1: Timeline of ‘Skill or Luck’ game (from perspective of agent i).

is exactly the more sophisticated version of the self-serving bias. The intuition why agents might furthermore overestimate the control they have over the outcome will be developed after a more formal analysis of the basic self-serving bias.

For the formal analysis, consider a game with timing as indicated in figure 1. After the agent has chosen which action to undertake, there is some probability $(1 - q)$ that luck intervenes, i.e. that the outcome will be determined by pure luck. In that case, the probability of success is p , independent of the action. In the other case, the outcome depends (probabilistically) on the action chosen by the agent. Success is then at least partially a matter of choosing good actions, i.e. a matter of ‘skill’.

An action’s probability of success (conditional on luck not intervening) is uncertain. Agents have subjective beliefs about this and these beliefs are commonly known. In particular, let p_n^i denote agent’s i ’s belief regarding the probability of success of action n (conditional on the outcome being determined by skill) and let them be i.i.d. draws from a distribution F on $[\underline{p}, \bar{p}] \subset [0, 1]$, i.e. $p_n^i \sim F[\underline{p}, \bar{p}]$ with $\underline{p} < \bar{p}$. Assume that F is atomless. Assume also that $\underline{p} < 1$ so that there is always some probability that luck fails.

Note, for further reference, the following more specific interpretation of this model. Let the projects come in good and bad types. As long as luck does not intervene, good projects always succeed, while bad projects always fail. Project types are unknown, but p_n^i is the belief of agent i that n is a good project. In this case, the outcomes in figure 1 when luck does not intervene could be relabelled ‘Good Project’ and ‘Bad Project’ instead of Success and Failure. (The outcomes when luck intervenes cannot be relabelled in that way since they are by assumption not related to the project type.)

Consider now the question of skill versus luck. Let $P_{j,A|S_i}$ denote the ex post probability according to agent j that a success of agent i was due to his action. Let $P_{A|S_i}$ denote the reference analogue. Say that ‘ i attributes success more to skill than j does’ if i ’s estimate of the probability that the outcome was determined by the action (rather than by pure luck) is

higher than j 's estimate, i.e. if $P_{i,A|S_i} \geq P_{j,A|S_i}$. Let PSB denote 'positive self-serving bias', so that $PSB_{i,j} = P_{i,A|S_i} - P_{j,A|S_i}$ and $PSB_i = P_{i,A|S_i} - P_{A|S_i}$ denote i 's 'positive self-serving bias' relative to respectively j and the reference belief. Analogously, let $NSB_{i,j} = P_{i,L|F_i} - P_{j,L|F_i}$ and $NSB_i = P_{i,L|F_i} - P_{L|F_i}$ denote the respective 'negative self-serving biases', and use SB when the property holds for both NSB and PSB. Remember, finally, that q denotes the ex-ante probability that the outcome is determined by the action and N the number of alternative actions.

Proposition 6 • $E[PSB_{i,j}] > 0$ for $N > 1$ while $PSB_{i,j} > 0$ a.s. as $N \rightarrow \infty$: On average (and a.s. as $N \rightarrow \infty$) an agent has a positive self-serving bias i.e. an agent who has a success attributes that success more to his skill than a randomly selected agent [would attribute the focal agent's success to skill]. Analogously, $E[NSB_{i,j}] > 0$ for $N > 1$ while $NSB_{i,j} > 0$ a.s. as $N \rightarrow \infty$: On average (and a.s. as $N \rightarrow \infty$) an agent attributes his failure more to pure (bad) luck than a randomly selected agent [attributes that failure to bad luck].

- Let A1 hold. $E[SB_i] > 0$ when $N > 1$ while $SB_i > 0$ a.s. as $N \rightarrow \infty$: On average (and a.s. as $N \rightarrow \infty$) and relative to the reference, a successful agent attributes his success too much to his own skill, while a failing agent attributes his failure too much to (bad) luck.
- Let $\bar{p} = 1$. In the limit as $N \rightarrow \infty$ $P_{i,L|F_i} \xrightarrow{a.s.} 1$, i.e. the agent is almost sure that any failure was due to bad luck. In case of a success, his belief that that was due to his own skill converges to $\frac{q}{q+p(1-q)}$.

Proof : Since $P_{j,A|S_i} = \frac{p_{Y_i}^j q}{p_{Y_i}^j q + p(1-q)}$ is strictly increasing in $p_{Y_i}^j$ and F^N FOSD F for $N > 1$, it follows that $E[P_{i,A|S_i}] > E[P_{j,A|S_i}]$. To show that $P_{i,A|S_i} > P_{j,A|S_i}$ a.s. as $N \rightarrow \infty$, it suffices to show that the probability that $p_{Y_i}^i > p_{Y_i}^j$ converges to 1. Note that $P[p_{Y_i}^i > p_{Y_i}^j] = \int P[x > p_{Y_i}^j] f_{p_{Y_i}^i}(x) dx = \int F(x) d[F^N(x)] = \frac{N}{N+1}$ which converges to 1 as $N \rightarrow \infty$. The arguments for failure are completely analogous. The second part of the proposition is also completely analogous. Consider now the final part of the proposition. An agent i who fails believes that his failure was due to (bad) luck with probability

$$\frac{(1-p)(1-q)}{(1-p)(1-q) + (1-p_{Y_i}^i)q} \xrightarrow{a.s.} \frac{(1-p)(1-q)}{(1-p)(1-q) + (1-\bar{p})q}$$

since $p_{Y_i}^i \xrightarrow{a.s.} \bar{p}$ as $N \rightarrow \infty$. This equals 1 when $\bar{p} = 1$. Analogously, the belief of a successful agent i that his success was due to skill is

$$\frac{p_{Y_i}^i q}{p_{Y_i}^i q + p(1-q)} \xrightarrow{a.s.} \frac{\bar{p}q}{\bar{p}q + p(1-q)}$$

which equals $\frac{q}{q+p(1-q)}$ when $\bar{p} = 1$. ■

This very last result might be a bit surprising. The reason for the difference between success and failures is as follows. When the agent has chosen a nearly perfect project, a failure becomes a nearly perfectly discriminating signal that luck intervened. On the upside, however, pure luck always has some chance of working out well. It follows that a success is never

perfectly discriminating and the agent can't exclude the possibility that the success was just due to luck. The expression reflects that intuition in more detail. As the probability of success by pure luck p decreases, a success must be more indicative of skill so that $P_{i,A|S_i}$ increases; As the prior probability that the outcome is due to skill $(1 - q)$ increases, so does the posterior probability $P_{i,A|S_i}$.

Comparative statics Consider now what happens when the number of available projects increases. To be precise, let a set of draws $\{p_n^i\}_{i=1,\dots,I;n=1,\dots,\infty}$ be given and compare the self-serving bias as the number of alternatives N (selected from this pool) varies.

Proposition 7 $P_{i,A|S_i}$, $P_{i,L|F_i}$, $E[PSB_{i,j}]$, $E[PSB_i]$, $E[NSB_{i,j}]$, and $E[NSB_i]$ all increase in N : For any given realization of beliefs, as the number of available projects increases, a successful agent is more likely to attribute his success to skill, a failing agent is more likely to attribute his failure to bad luck, and all self-serving biases increase. When $N = 1$ and $p_n \sim F$, $E[PSB_{i,j}] = E[PSB_i] = E[NSB_{i,j}] = E[NSB_i] = 0$.

Proof : From the proof of proposition 6, we know that $P_{i,A|S_i}$ increases in $p_{Y_i}^i$. Since $Y_i = \arg\max_{Y \in \mathbf{A}} p_Y^i$, $p_{Y_i}^i$ weakly increases when we add an extra project to \mathbf{A} (while keeping the beliefs in the other projects unchanged). The first part of the proposition then follows immediately, while the second part is analogous. That $E[PSB_{i,j}]$ increases in N follows from the fact that F^{N+1} FOSD F^N so that $E[P_{i,A|S_i}]$ increases while $E[P_{j,A|S_i}]$ remains unchanged. The other arguments are analogous. Finally, when $N = 1$ and $p_n \sim F$, we have that $E[P_{i,A|S_i}] = E[P_{j,A|S_i}] = E[P_{A|S_i}]$ since they all have identical distributions. ■

The intuition is that, from the focal agent's perspective, more available actions can only improve $p_{Y_i}^i$, the subjective probability of success of the action he ends up undertaking, since it equals $\max_{n \in N} p_n^i$. While this result holds for the expected value, it is not necessarily true for any specific realization. The reason is that when the focal agent chooses the newly offered project since it has higher probability of success than any he had already in hand, we also get a new draw for the 'observer' and that draw may be a lot higher than for the original project. Note also that in the other extreme that there is only one possible task, the biases disappear in expectation.

This comparative static might actually be quite important. In particular, as mentioned earlier, experiments that test for the self-serving bias typically restrict the actions the subjects can take. In everyday life, people have much more freedom. This implies that such structured experiments will tend to under-estimate the practical relevance of these biases.

A second comparative static concerns the form of the distribution⁹. The main conclusion is that the self-serving bias increases in the underlying uncertainty, but the result is again much broader than that and can be related to moments of any order, including skewness and curtosis.

Proposition 8 Let F and G be analytic functions.

- If G is a mean-preserving spread of F , then $\exists \tilde{N}$ s.t. $\forall N \geq \tilde{N} : E_G[SB_{i,j}] > E_F[SB_{i,j}]$
: In expectation the level of self-serving bias increases in the underlying uncertainty

⁹See section 4 for the definitions of mean-preserving spread and moment-preserving stochastic dominance.

- If F MP SD_m G for some even $m \geq 2$, then $\exists \tilde{N} s.t. \forall N \geq \tilde{N}, E_G[SB_{i,j}] > E_F[SB_{i,j}]$.
With m odd, the inequalities go the other way.

Robustness As before, it is useful to ask how far the results are robust to changes in the specification. In particular, while the game structure in the above analysis captures quite well the idea of ‘skill versus luck’ and lends itself directly to an experimental implementation, it is not very general. The following would be more general. Let $P(p, l)$ denote the overall probability of success of the agent’s action and let it depend on two factors which correspond to the two arguments p and l . On the one hand, the final probability of success P depends on the agent’s choice of action. In particular, the argument p is completely determined by the action choice, with action n resulting in $p = p_n$. The value p_n of a particular action is unknown, but each agent has a subjective belief. Let agent i believe that for action n , $p = p_n^i$ with probability 1. On the other hand, the probability of success P is also influenced by some random exogenous factors which are captured by the second argument l . In particular, let l have a prior distribution σ . Assume that P increases in both its arguments.

Upon observing the outcome, agent i can form an ex-post belief about the value of the exogenous influence l . A higher ex-post assessment of l , in the sense of a FOSD-shift of the ex-post distribution, corresponds to attributing the success more to ‘luck’ and thus relatively less to ‘skill’, in the sense of a skillful choice of action.

The original game can be reformulated to fit this framework. In particular let $l \in \{L, M, H\}$, which we assume to be completely ordered from left to right, and

$$P = \begin{cases} 1 & \text{if } l = H \\ p_n^i & \text{if } l = M \\ 0 & \text{if } l = L \end{cases}$$

Conditional on success, it must be that $l \in \{M, H\}$. The ex-post distribution of l is thus completely characterized by the probability that $l = H$, which indeed equals the ex-post probability that the outcome was determined by pure luck in the original setup.

Going back to the more general case, the following proposition shows that the result is quite robust to this more general specification. Let l be ex-ante uniformly distributed on the interval $[0, 1]$ or have, conditional on success, a discrete distribution with only two possible values. Let $\sigma(l | S)$ denote the ex-post belief on l conditional on a success.

Proposition 9 • If $\log P(p, l)$ is submodular then an increase in the argument p causes an inverse FOSD shift of $\sigma(l | S)$. In other words, a subjectively better choice of action decreases the (subjective) ex-post assessment of the importance of luck in a success.

- If, for a given \hat{p} , \hat{l} is such that $P(\hat{p}, l) = 1 \forall l \geq \hat{l}$ then $P[l \geq \hat{l} | S]$ decreases in \hat{p} (at $p = \hat{p}$).

While the first part of the proposition specifies only a sufficient condition, it shows already that the result is quite robust. The following corollary shows indeed that the most obvious specifications fall in that category.

Corollary 2 *When $P(p, l)$ is submodular, or $P(p, l) = f(p) \cdot g(l)$ or $P(p, l) = f(p) + g(l)$ with both f and g increasing, then an increase in p causes an inverse FOSD shift of $\sigma(l | S)$.*

Since submodularity of P is sufficient, the result holds as long as the arguments not too strong complements. This submodularity corresponds to the idea of ‘skill or luck’. The second part of proposition 9 might seem a bit more obscure. It essentially says that, for any distribution, ‘choosing a more attractive action causes success to be less attributed to extreme luck’, where ‘extreme luck’ means that the outcome is always a success, independent of the action. Note that this corresponds exactly to what happened in the original skill or luck game.

This robustness discussion also points to a very different issue. A typical experimental setup, used in e.g. Streufert and Streufert (1969) and Wolosin, Sherman, and Till (1973), would ask the subjects to allocate 100 points to their own skill and to situational circumstances. With the above probability structure, however, there are many ways how such allocation could be done. This is a very general problem. The question ‘in how far do you attribute this to skill versus luck’ is ambiguous. It is not surprising then that such relatively unstructured experiments sometimes give inconsistent results.

Except for being more transparent, the original skill or luck model also has the advantage that we can implement it nearly directly as an experiment.

5.1 The failure to learn from failures

From an average person’s perspective, a person with extreme beliefs rejects too easily disconfirming evidence as a random variation. While this might look irrational, the previous analysis shows that that is not necessarily the case, depending on the beliefs of the focal person. People with differing beliefs ‘see’ different things in the same observations. This fits the experimental evidence: Lord, Ross, and Lepper (1979), for example, showed how both proponents and opponents of the death penalty interpreted identical evidence as bolstering their case. Given that they see different things, people with differing beliefs therefore also disagree on the amount of updating that should take place. In particular, in a world with differing priors, rational people generally don’t learn as much from their own failures as they should from the perspective of a rational outside observer.

To see this formally, consider again the earlier mentioned interpretation of the model. A project is either good or bad. As long as luck does not intervene, good projects succeed while bad projects fail. (If luck intervenes then the outcome is determined independently of the type of the project.) The type of a project is unknown but agent i thinks that project n is good with probability p_n^i . As before, the ‘reference’ agent also does not know the project’s type, but has a belief p_n .

Assume now that agent i chose an action, say \hat{n} , which resulted in a failure. Let $P_{L|F_i}$ denote the ‘reference’ ex-post probability that i ’s failure was due to bad luck. Let further \tilde{p}_n^i denote i ’s updated belief about n if he knew the reference $P_{L|F_i}$ (but still started updating from his own prior p_n^i):

$$\tilde{p}_n^i = 0 \cdot (1 - P_{L|F_i}) + p_n^i P_{L|F_i} = p_n^i P_{L|F_i}$$

and let \hat{p}_n^i denote his effective updated belief:

$$\hat{p}_n^i = 0 \cdot (1 - P_{i,L|F_i}) + p_n^i P_{i,L|F_i} = p_n^i P_{i,L|F_i}$$

From the reference perspective, the agent does not learn enough from his own failure:

Corollary 3 *Under A1, $\hat{p}_n^i > \tilde{p}_n^i$ a.s. as the number of alternative actions $N \rightarrow \infty$: As $N \rightarrow \infty$, the agent almost surely learns less from his own failure than he should from the reference perspective.*

Proof : This follows immediately from the above equations and the fact that by proposition 6 $P_{i,L|F_i} > P_{L|F_i}$ a.s. when $N \rightarrow \infty$. ■

Note that, from a subjective perspective, agents learn precisely what they should learn, since they are Bayesian rational. However, from a ‘reference’ perspective, they discard too much negative evidence as ‘random variation’. Agents do learn, however, and with sufficient feedback the beliefs will converge. This theory therefore does not allow to explain the experimentally documented fact that beliefs may sometimes diverge more when additional evidence is presented.

There are other interesting interactions with learning. It has, for example, been suggested that there is also a sense in which this mechanism may actually accelerate learning: people tend to undertake, and thus learn about, precisely these actions about which they are most out of line with others. More in general, the rate of learning and its comparative statics seems to be a very interesting line of research.

5.2 An illusion of control

This subsection now shows how agents might come to over-estimate the control they have over the outcome. To that purpose, consider the following modification of the basic setting. Let the probability that luck intervenes vary by action. Define the ‘level of control’ to be the probability that the outcome depends on the agent’s action. Remember that the level of control was originally the same for all actions and denoted q . In what follows, different actions will have different levels of control associated with them. The level of control is unknown, but agents have subjective beliefs on this variable.

To see the intuition why agents might come to over-estimate the control they have, note the following. Agents who believe their choice of action has a higher likelihood of success than pure luck will want to select actions that give them maximal control, i.e. that maximize the probability that the outcome is determined by their action instead of by pure luck. But with differing priors on the level of control, an agent will again tend to select actions on which he over-estimates his control from the perspective of a randomly selected outsider. This is the ‘random variation plus systematic choice gives systematic variation’ effect mentioned earlier.

Notice further that, according to equation (1), an agent will attribute success more to skill as the prior probability that the outcome is pure luck decreases. The combination of these effects will cause an asymmetry in the self-serving bias: the agents’ self-enhancing bias will be stronger than their self-protective bias.

For the formal analysis, consider a situation in which there is not only uncertainty about the probability of success of the actions, p_n^i , but also about the probability that luck will intervene on some particular course of action. In particular, let q_n^i denote agent's i 's prior belief that the outcome of action n is determined by skill and let it be drawn from a distribution G on $[\underline{q}, \bar{q}] \subset [0, 1]$, i.e. $q_n^i \sim G[\underline{q}, \bar{q}]$ with $\underline{q} < \bar{q}$. The agent thus believes that if action n is undertaken, the outcome is determined by pure luck with probability $1 - q_n^i$. Let G have no atoms. Assume finally that $p < \bar{p}$ so that there is at least some positive probability that actions outperform pure luck.

The expected payoff of action n from the perspective of agent i is

$$q_n^i p_n^i + (1 - q_n^i)p = p + q_n^i(p_n^i - p)$$

where p is still the probability of success when luck determines the outcome and p_n^i denotes agent i 's belief on the probability of success of action n (conditional on luck not intervening). It follows that whenever possible, the agent will choose an action with $p_n^i > p$, i.e. an action that has a higher probability of success than luck. Moreover for every action with $p_n^i > p$, the agent prefers higher q_n^i . This is intuitive: the agent wants maximal control if he thinks he will do better than pure luck. This further suggests that with large enough choice, the agent will end up selecting an action with q_n^i and p_n^i near their respective maxima. That, and its consequences, is essentially the content of the following proposition¹⁰.

Proposition 10 *As $N \rightarrow \infty$*

- $q_{Y_i}^i > q_{\bar{Y}_i}^i$ a.s. : *From the perspective of any outsider, i almost surely over-estimates his control.*
- *If $\bar{q} = 1$ and $\bar{p} = 1$ then $P[\text{Skill} \mid \text{Success}] \rightarrow 1$, while $P[\text{Luck} \mid \text{Failure}]$ depends on the tails of the distributions of p_n^i and q_n^i : Upon success the agent is almost sure the success is due to his skillful choice of action while his inference from a failure can go either way depending on the specifics of the situation.*

5.3 Comparison with the experimental evidence

The mechanism presented here fits the basic experimental evidence in the sense that it consistently leads to a self-enhancing bias and typically leads to a self-protective bias, though the latter is sensitive to the context. While absent from the current model, it might also help explain why high self-esteem leads to more self-serving bias¹¹.

There are, however, some important issues and caveats. First of all and mentioned earlier, the majority of the existing experimental evidence relies heavily on the use of anagrams, and

¹⁰ Y_i again denotes agent i 's preferred action.

¹¹In particular, agents who consider themselves better at choosing actions will have more extreme estimates of the probabilities of success of alternative actions and thus relatively higher estimates of their likelihood of success. A second interpretation is that they consider themselves better at implementing actions, which causes a FOSD shift in the probability distribution and thus also a higher estimate of the likelihood of success. The latter explanation is independent of the current model, however.

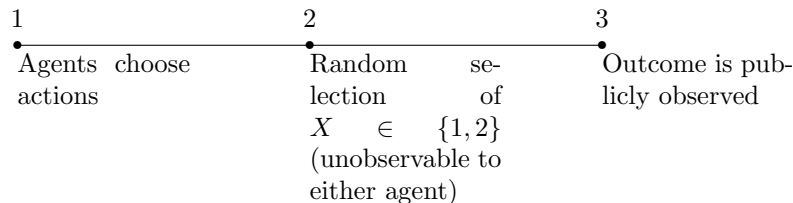


Figure 2: Timeline of ‘Praise or Blame’ game

thus does not correspond very well with the issues addressed in the model. It is, on the other hand, comforting that a very casual study of the results of Zuckerman (1979) suggests that the evidence for the self-serving bias is stronger for the experimental designs that fit the current model better. Second, there does not seem to be any experimental studies that allow to assess the validity of the comparative statics derived above. Finally, there are also experimental results that cannot be explained with the current model. For example, the results that the self-serving bias is influenced by gender or by the expectation that one will perform the same test again in the future cannot be readily explained by the current model.

The experimental literature on the illusion of control has focused on purely random gambling situations. If we accept that agents in these experiments actually believe that different action really *do* have different probabilities of success, then the current model fits the evidence. If, on the contrary, we completely reject that possibility then there is to my knowledge no existing evidence that allows us to assess the illusion of control predictions derived here.

6 Praise yourself, blame the others

This section considers the implications of these mechanisms for cooperative ventures. It shows how the agents’ self-estimated contributions to cooperative projects may add up to more than 100% when the project turns out to be a success.

The model is based on the model of section 5. Only now, there are two agents and the uncertainty is about whose action will determine the outcome, rather than whether the outcome was due to skill or luck. We study how responsibility for success and for failure is ex-post attributed by the agents.

Let there be two individuals, denoted 1 and 2, who each publicly and simultaneously choose an action. Let there be N potential actions A_n with $1 \leq n \leq N$. The overall outcome can be either a success or a failure, denoted S and F respectively. The timing is according to figure 2. After each agent has chosen what activity to undertake, it is determined randomly and unobservably whose action will determine the outcome. In particular, the random variable $X \in \{1, 2\}$ indicates whether it is agent 1’s or agent 2’s action that determines the outcome. Let $P[X = 1] = P[X = 2] = 1/2$. The probability of success of an action is not known, but both agents have beliefs about that probability and these beliefs are common knowledge. In particular, let agent’s i ’s belief on the probability of success of action n be

denoted p_n^i and let it be drawn from a distribution F on $[\underline{p}, \bar{p}]$, i.e. $p_n^i \sim F[\underline{p}, \bar{p}]$ with $\underline{p} < \bar{p}$.

We are interested to see how agents make inferences about who is responsible for the outcome, based on whether the outcome is a success or a failure. Let $\tau_i^S = \nu_i(X = i \mid S)$ denote agent i 's belief about the probability that he was responsible for the outcome conditional on success and $\tau_i^F = \nu_i(X = i \mid F)$ analogously for failure. The following proposition says that the self-estimated contributions tend to add up to more than 100% when the outcome is a success and to less than 100% when it is a failure. Moreover, the discrepancy increases in the number of alternative projects. This points again to the importance of choice for this result.

Proposition 11 *As the number of alternative actions $N \rightarrow \infty$,*

- $\forall N > 1$, $E[\tau_1^S + \tau_2^S] > 1$ and increases in N , while $E[\tau_1^F + \tau_2^F] < 1$ and decreases in N .
- As $N \rightarrow \infty$, $\tau_1^S + \tau_2^S > 1$ a.s. and $\tau_1^F + \tau_2^F < 1$ a.s..

To see how substantial the effect can be, we can calculate the expected sum of (subjectively estimated) contributions. For example, when $p_n^i \sim U[0, 1]$, $E[\tau_1^S + \tau_2^S] \xrightarrow{\text{a.s.}} 2 \ln(2) \approx 1.4$.

Comparison with the experimental evidence As before, the conclusions of the model are consistent with the general evidence. The model does not, however, explain the result in Sedikides, Campbell, and Reeder (1998) that agents who have a close personal relationship suffer less from the egocentric bias. There is also no evidence that throws light on the validity of the comparative statics.

It should also be noted that although the famous study by Ross and Sicoly (1979) does not fit the above model (since there was no success or failure in their experiment), it *is* nevertheless closely related. In particular, a variation on the model of section 7 presented in Van den Steen (2001) does fit that experiment and its results.

7 The Lake Wobegon Effect¹²

Research shows that more than 80% of US drivers consider themselves above the median (Svenson 1981). In a different study, less than 5% of employees rated themselves below the median (Meyer 1975). Section 4 on overoptimism implicitly hinted at explanations for these phenomena. This section presents a slightly modified version of the basic model that fits the phenomena better.

To fix ideas, consider the example how drivers rate themselves. Assume that people agree that driver quality refers to the probability of causing an accident over the next 100,000 miles of driving, but that they have differing beliefs as to what it takes to be a good driver, i.e. to have a low risk to cause an accident. Assume further that people care about how well they

¹²This terminology refers to Garrison Keillor's radio sign-off from fictional Lake Wobegon, Minnesota, 'where the women are strong, the men are good-looking, and all of the children are above average'. (It was pointed out to me by John Roberts.)

drive. Then all drivers will act differently, each following his own beliefs about safe driving. But each driver will also evaluate his own and others' driving styles with the criteria that he used to choose his actions. This double use of criteria favors his own driving style. This mechanism is identical to the earlier overoptimism effect when you look from one agent's perspective to the own and others' actions.

To see the current result more formally, let driver quality q , as measured by the probability of causing an accident over the next 100,000 miles, be a function of two factors :

1. 'how well you anticipate dangerous situations', denoted by s for 'situations', and
2. 'how well you obey the traffic rules', denoted by r for 'rules'.

Let r and s depend on the amount of attention paid to each, denoted respectively by $a_r \geq 0$ and $a_s \geq 0$. In particular, let driver quality be a function

$$q(\alpha, a_r, a_s) = \alpha r(a_r) + (1 - \alpha)s(a_s)$$

where $\alpha \in [0, 1]$ denotes the relative importance of following the traffic rules. Assume all functions to be smooth, r and s to be strictly increasing and strictly concave in attention, with $r'(0) = \infty$ and $s'(0) = \infty$. Let the total amount of attention that a driver can spend be \bar{a} , so that $a_s = \bar{a} - a_r$. An agent thus chooses a_r subject to $0 \leq a_r \leq \bar{a}$, which gives him a continuum of potential actions. Note that if we allow the number of actions in section 4 to go to infinity, then the probability that two agents undertake the same action goes to zero so that each agent almost surely thinks that he is strictly better than all other agents. The current model will come to an analogous conclusion for driving quality.

To continue with the model, assume that each person cares about his quality as a driver. In particular, person i 's utility function u_i is

$$u_i = E_\alpha^i[1 - q(\alpha, a_{r,i}, a_{s,i})] = 1 - q(\alpha_i, a_{r,i}, a_{s,i})$$

where $a_{r,i}$ and $a_{s,i}$ denote i 's choice of attention levels, and α_i denotes the mean of i 's belief about the importance of following traffic rules. Note that q is linear in α so that only α_i , the expected value of α , matters for i 's expected utility.

Consider now I agents, each with his own prior belief about the expected value of α , i.e. with his own individual belief on the relative importance of r and s . In particular, let the mean of each agent's belief be independently drawn from a distribution F with support contained in $[0, 1]$ and with continuous density. Let that belief be known to all other agents.

This simple setting gives the following result :

Proposition 12 *Upon observing everyone else's driving behavior, each agent concludes that he/she is (strictly) the best driver of the whole population.*

There are essentially 3 elements in this model and the one of section 4 which combine to give the inconsistent assessments :

1. People care about how well they perform.

2. The performance is affected by actions and decisions under the agent's control.
3. There is disagreement what actions will result in good performance.

Note in particular that any evaluation inconsistencies disappear in the current model if we look at how people rate themselves on a component, say 'anticipating dangerous situations', rather than on the aggregate behavior. This fits with the empirical evidence (Allison, Messick, and Goethals 1989, Wade-Benzoni, Thompson, and Bazerman 2000).

In this analysis every agent knows that all other agents also consider themselves better than the median and understands why. While this mutual knowledge and understanding is not necessary, it shows explicitly that part of this bias may persist even when people are made aware of the others' self-evaluation.

8 Organizational implications

These biases have important implications for firms and organizations. Larwood and Whitaker (1977), for example, showed how managers overestimate growth prospects. Furthermore, annual reports also display self-serving biases in their causal attributions (Bettman and Weitz 1983). This suggests that firms might not learn enough from their strategic and other failures. The rest of this section explores some potential implications of these biases for organizations.

Evaluations and incentives Baker, Jensen, and Murphy (1988) discuss the impact of the self-enhancing bias (section 7) on the effectiveness of incentive systems. In particular, they argue that the bias may explain why superiors are reluctant to give their employees bad performance evaluations: the latter would consider such evaluations unfair and become dissatisfied. They further argue that '[b]iased (...) performance evaluation reduces productivity by reducing the effectiveness of incentives in the organization.'

The current analysis suggests that this conclusion is not necessarily warranted. First of all, it assumes that superiors give good evaluations to avoid the conflict that arises from employees' *dissatisfaction* with their rating (Baker, Jensen, and Murphy 1988). Equity theory (Adams 1965), however, suggests that the perception of fairness is also an important determinant of *motivation*. This psychological theory, which has been confirmed experimentally (Goodman and Friedman 1971, Miles, Hatfield, and Huseman 1989), holds that people try to make sure the balance between efforts (input) and what they get from that effort (outcome) is fair. For example, people who feel they get underpaid will reduce their effort. Compressing evaluations on the better half of the scale might thus actually *improve* the effectiveness of the incentive system: the evaluations seem fair since they match employees' perceptions and such fairness improves motivation. And motivation is what incentives are about. Second, an employer can anticipate his superiors' tendency to avoid negative evaluations and adapt nominal bonuses and incentives to compensate for it. In particular, the nominal bonuses should get steeper and lower than they would be without the bias. This modification may sometimes completely restore the effectiveness of the incentives.

Not all problems will be eliminated, however. For example, the effectiveness of the evaluation system in terms of providing feedback will still be reduced since the feedback system now makes the same ‘errors’ as the agents made in section 5.1 on learning. The threat of layoffs as an incentive mechanism is also less effective since an employee with ‘above average’ ratings will generally not consider himself at risk. For these reasons it is sometimes important to consider remedies. The analysis in section 7 showed that the inconsistency is driven by disagreement on the relative importance of different behaviors. A possible solution is to evaluate the behaviors separately and to be explicit on how the results get aggregated to an overall evaluation.

Decision processes Kaplan (1986) and Poterba and Summers (1995) decried the fact that companies often use too high hurdle rates in Discounted Cash-Flow Analysis. Doing so reduces the attractiveness of projects with cash-flows that come far in the future. Their arguments make perfect sense if managers can form an unbiased estimate of the future cash flows of suggested projects. The analysis here shows, however, that managers will tend to overestimate the return on their projects, and more so as there is more uncertainty. It is furthermore reasonable to assume that there is more uncertainty about cash-flows as they are further in the future. The use of artificially high discount rates, while imperfect, might be a practical way to compensate for this tendency. For the same reasons, it also makes sense to discount benefits that are difficult to quantify.

The analysis also has potential implications for the theory of delegation. In particular, it suggests that delegation to an employee might improve efficiency if that employee under-supplies effort and effort is complementary to the probability of success. Along the same vein, centralization of the decision should be optimal if effort is a substitute to the probability of success. This is similar to Zabochnik (2001).

Organizational learning Section 5.1 shows how the self-serving bias might inhibit learning from failure or mistakes (and make decision makers overinterpret the significance of their successes). In cases where learning is important, organizations should look for ways to eliminate these effects.

One way to improve learning is to let someone other than the original decision maker review the failures. Such independent observer does not have the endogenous overoptimism and will therefore tend to be more accurate in his or her evaluation. On an organizational level, this suggests for example that quality control should not be part of the manufacturing department, as indeed specified by the ISO-9000 certification and the Baldrige Awards, and suggests a function for independent directors. It also provide a rationale beyond mere expertise for the existence of a company like Exponent, Inc. (formerly ‘The Failure Group, Inc.’) which specializes in general failure analysis. A further example are the FAA reviews of incidents and accidents in air travel and transportation. Note, however, that this paper also suggests that such outside reviewer will generally be at odds with the original decision maker on the causes of the failure. This is consistent with the fact that the pilot union often rejects the conclusion of the FAA that an accident was due to pilot error. The resentment that such arguments can generate and its impact on performance should be considered before taking

this approach. Moreover, such independent reviews will only work if that outsider has the means to influence the future actions of the decision maker. In the absence of such influence, it might be necessary to replace the original decision maker. This is a very different theory for CEO replacement than the idea that managers get fired because they are not good enough or did not spend enough effort.

Bargaining and litigation The implications of the self-serving bias for bargaining and litigation have been explored in some detail (Priest and Klein 1984, Babcock and Loewenstein 1997, Babcock and Loewenstein 1997). It has, in particular, been argued that such bias increases the probability of breakdowns in bargaining, of strikes in labor negotiations and of effective litigation. The experimental evidence in this field largely supports the claims made. These experiments, however, usually do not endogenize the actions that lead to bargaining and litigation. Instead, the experimental subjects get to read a description of the case before or after they have been told which role they will play in the case. This setup excludes essentially the mechanisms that have been discussed here. This analysis thus suggests that these experimental results may underestimate the magnitude of the real self-serving bias and thus its role in real-life strikes, litigation and bargaining breakdowns.

9 Conclusion

The key conclusion of this paper is that as soon as we relax the assumption that people somehow happen to have identical and true priors, a number of well-established biases surface very naturally. This paper thus shows that even rational agents may suffer from these attribution and inference biases. It indicates furthermore when such biases are most likely to appear and takes some steps in studying implications and potential remedies. An important implication of the paper is that such biases are partially endogenous. This might affect the conclusions of studies that consider their implications for e.g. market outcomes.

While the theory is largely consistent with the existing empirical evidence, it is important to test it explicitly as an alternative hypothesis. Experimental economics seems to be the most appropriate methodology. In particular, one of the key advantages of the models in this paper is that they lend themselves directly to an experimental implementation. While the comparative statics could be a possible starting point, it seems that it would be more effective to directly test the basic mechanism. The theory might also be tested empirically by looking for example at the behavior and payoff of gamblers.

The paper also suggests further areas of research. On the more theoretical side, I mentioned already the issue of learning under differing priors and how this may be affected by the ideas presented in this paper. The section on organizational implications also suggested some more applied research areas. The most interesting of these are probably the implications for incentives and for decision making processes.

Most importantly, the paper suggests the value of studying the implications of differing priors. It seems, in particular, that some behaviors that are traditionally considered to be in the realm of bounded rationality might actually be usefully studied by using differing priors.

A Moment-Preserving Stochastic Dominance

Let F and G have support contained in $[0, 1]$. Define $F_0(x) = F(x)$ and $F_k(x) = \int_0^x F_{k-1}(t)dt$. Note that, if F is atomless, then F is continuous, which implies that F_k is k -times continuously differentiable. Say now that F ‘lower moment preserving m -order stochastically dominates’ G , denoted $F \text{ MPSPD}_m G$ if

1. $F_k(1) = G_k(1)$ for $k = 0, \dots, m-1$
2. $F_{m-1}(x) \leq G_{m-1}(x)$ with strict inequality for at least one x

Note that the latter condition implies that $F_m(1) < G_m(1)$.

Proposition 1 *Let $m \geq 1$. If $F \text{ MPSPD}_m G$, then*

1. *they have identical moments up to order $m-1$*
2. *F ’s m ’th moment is larger than G ’s when m is odd and smaller when it is even.*

Proof : By repeated integration by parts, we can write for any $k \leq m$

$$\begin{aligned}
 & \int_0^1 x^k (f - g) dx \\
 &= [x^k [F(x) - G(x)]]_0^1 - k \int_0^1 x^{k-1} [F(x) - G(x)] dx \\
 &= [x^k [F(x) - G(x)]]_0^1 - k [x^{k-1} [F_1(x) - G_1(x)]]_0^1 + k(k-1) \int_0^1 x^{k-2} [F_1(x) - G_1(x)] dx \\
 &= \sum_{l=0}^k \left[(-1)^l \frac{k!}{(k-l)!} x^{k-l} [F_l(x) - G_l(x)] \right]_0^1 \\
 &= (-1)^k k! [F_k(1) - G_k(1)]
 \end{aligned}$$

When $k < m$ this gives $\int_0^1 x^k (f - g) dx = 0$ while it gives $(-1)^m \int_0^1 x^k (f - g) dx = k! [F_m(1) - G_m(1)] < 0$ when $k = m$ since when $F \text{ MPSPD}_m G$ then $F_m(1) < G_m(1)$. This proves the proposition. \blacksquare

Proposition 2 *If $u(x)$ is a smooth function with the k ’th derivative $(-1)^k u^{(k)} < 0$ for $k > 0$ and $F \text{ MPSPD}_m G$, then $\int_0^1 u(x) [F(x) - G(x)] \geq 0$.*

Proof :

$$\begin{aligned}
 & \int_0^1 u(x) [F(x) - G(x)] \\
 &= [u(x) [F_1(x) - G_1(x)]]_0^1 - \int_0^1 u'(x) [F_1(x) - G_1(x)] dx \\
 &= - \int_0^1 u'(x) [F_1(x) - G_1(x)] dx \\
 &= \dots \\
 &= \left[(-1)^{m-2} u^{(m-2)}(x) [F_{m-1}(x) - G_{m-1}(x)] \right]_0^1 + (-1)^{m-1} \int_0^1 u^{(m-1)}(x) [F_{m-1}(x) - G_{m-1}(x)] dx \\
 &= \int_0^1 (-1)^{m-1} u^{(m-1)}(x) [F_{m-1}(x) - G_{m-1}(x)] dx > 0
 \end{aligned}$$

\blacksquare

B Alternative Specification of Belief Distributions

This appendix shows that the results of the main paper still hold when the subjective beliefs are derived as deviations from the reference belief. In particular, make the following assumption.

Assumption 2 *The reference belief p_n is i.i.d. (over projects) distributed according to some distribution H . The subjective belief of agent i is $p_n^i = p_n + \epsilon_{i,n}$ with $\epsilon_{i,n}$ i.i.d. distributed according to some distribution $K_{(p_n)}$, which may depend on the realization of p_n , and has mean zero and a support with non-empty interior.*

Note that I allow K to depend on the realization of p_n . Note also that K having mean zero and non-empty interior implies that the maximum of the support of H must be strictly smaller than 1. Note finally that in this case the agents' beliefs are not any more independently distributed (as I assume for the subjective arguments). The assumption does, however, satisfy the conditions of proposition 3.

Under this assumption, the results of the main paper continue to hold.

Proposition 3 *Under A2, $E[\rho_i] > 0$ and $E[\rho_i^j] > 0$ when $N > 1$ and $\rho_i > 0$ and $\rho_i^j > 0$ a.s. as $N \rightarrow \infty$. In expectation (and a.s. when $N \rightarrow \infty$) the agent is overoptimistic about the likelihood of success of her project.*

Proof : For the first part of the proposition, fix some agent i . Let B_n denote the event ' $p_n + \epsilon_{i,n} = \max_{k \in N} p_k + \epsilon_{i,k}$ '. This gives $E[\rho_i] = \int \cdots \int \sum_{n \in N} \epsilon_{i,n} I_{\{B_n\}} dF$. It thus suffices to show that the expected value of $\epsilon_{i,n}$ conditional on B_n , is positive. To that purpose, condition on B_n , on the realizations of all N reference probabilities $\{p_1, \dots, p_N\}$ and on the realizations of all $\epsilon_{i,k}$, except for $\epsilon_{i,n}$. The conditional expectation of ρ_i is $\int \epsilon_{i,n} I_{\{\epsilon_{i,n} \geq \max_{k \neq n} p_k + \epsilon_{i,k} - p_n\}} dK_{p_n}$ which, given that $\epsilon_{i,n}$ is mean-zero distributed, is always non-negative and at least for some cases strictly positive. This implies the first part of the proposition under A2. The second part of the result under A2 follows from proposition 3. ■

C Proofs

C.1 Endogenous Overoptimism

For all the proofs that follow, note that the distribution of $p_{Y_i}^j$ is F , while that of $p_{Y_i}^i$ is F^N . Moreover, as $N \rightarrow \infty$, $p_{Y_i}^i \xrightarrow{\text{a.s.}} \bar{p}$. For what follows, consider two distribution functions, F and G , that are analytic on $[0, 1]$. If $\exists x : F(x) > G(x)$, define $\bar{x}_m = \sup(x : (-1)^m[F(x) - G(x)] > 0)$, else let $\bar{x}_m = 0$. Define $\bar{x} = \bar{x}_0$. Say as before that F MPSPD $_m$ G if $F_k(1) = G_k(1)$ for $k = 0, \dots, m-1$, and $F_{m-1}(x) - G_{m-1}(x) \leq 0$ with strict inequality for at least one x .

Proof of Proposition 2b: If F MPSPD $_2$ G then $F_1(\bar{p}) - G_1(\bar{p}) = 0$ and $F_1(x) - G_1(x) \leq 0 \forall x \in [\bar{p}, \bar{p}]$ with strict inequality for some. We need to show that $\left[1 - \int_0^1 F^N dx\right] - \left[1 - \int_0^1 G^N dx\right] < 0$ or $\int_0^1 F^N dx > \int_0^1 G^N dx$. For any k such that $0 \leq k \leq N-1$, integration by parts shows:

$$\begin{aligned} & \int_0^1 (F - G) F^{N-k-1} G^k dx \\ &= \left[F^{N-k-1} G^k \int_0^y (F - G) dx \right]_0^1 - \int_0^1 \int_0^y \{(F - G) dx\} d(F^{N-k-1} G^k) \\ &= - \int_0^1 \int_0^y \{(F - G) dx\} d(F^{N-k-1} G^k) = \int_0^1 (G_1 - F_1) d(F^{N-k-1} G^k) \\ &> 0 \end{aligned}$$

by $\int_0^1 (F - G) dx = 0$, the definition of SOSD, and the fact that $F^{N-k-1} G^k$ is just a distribution function with support contained in $[0, 1]$. This proves the first part of the proposition.

The first part of the second bullet follows lemma 1 with $z(x) = x$.

For the second part of the second bullet, note that $p_{Y_i}^i \xrightarrow{\text{a.s.}} \bar{p}$ and that, by lemma 3, \bar{p} weakly decreases for a MPSPD $_m$ shifts with m even. ■

Lemma 1 *If F and G are analytic functions and F MPSPD $_m$ G for some even $m \geq 2$, then $\exists \tilde{N}$ s.t. $\forall N \geq \tilde{N}$, $E_{F^N}[z(x)] < E_{G^N}[z(x)]$ for any $z(x)$ that is strictly increasing and continuously differentiable on $[0, 1]$. When m is odd, the opposite inequalities hold.*

Proof : This follows as follows from the following lemma's. Let m be even. According to lemma 3 $\exists \underline{x} < \bar{x}_m$ s.t. $\forall x \geq \underline{x}$, $(F - G) \geq 0$ and for $\underline{x} \leq x < \bar{x}_m$, $[F(x) - G(x)] > 0$. By lemma 4 then, $\exists N$ such that $\forall n \geq N$, $E_{F^n}[z(x)] < E_{G^n}[z(x)]$ for any $z(x)$ that is strictly increasing and continuously differentiable on $[0, 1]$. The argument for m odd is analogous. ■

Lemma 2 *Let H be analytic on $[0, 1]$, $H(0) = 0$, $H(x) \geq 0$ on $[0, \bar{x})$ with strict inequality for some x , and let $\underline{x} = \inf(x > 0 : H(x) > 0)$. $\exists \bar{x}_1 > \underline{x}$ such that $H'(x) = h(x) \geq 0 \forall x \in [0, \bar{x}_1)$ and $H'(x) = h(x) > 0$ for all $x \in (\underline{x}, \bar{x}_1)$.*

Proof : Since h is analytic, its Taylor expansion implies that for $x \in (\underline{x}, \bar{x})$,

$$h(x) = \sum_{n=0}^{\infty} \frac{(x - \underline{x})^n}{n!} h^{(n)}(\underline{x})$$

Note that it cannot be that $h^{(n)}(\underline{x}) = 0$ for all $n \geq 0$ since we would then have (by the Taylor expansion) $H(x) = \sum_{n=0}^{\infty} \frac{(x - \underline{x})^n}{n!} H^{(n)}(\underline{x}) = H(\underline{x}) = 0$ in $x \in (\underline{x}, \bar{x})$, where $H(\underline{x}) = 0$ follows from the definition of \underline{x} and

continuity of H . So there exists some smallest n such that $h^{(n)}(\underline{x}) \neq 0$. Let that be \hat{n} . We now claim it must be that $h^{(\hat{n})}(\underline{x}) > 0$. Assume to the contrary that $h^{(\hat{n})}(\underline{x}) < 0$. Note that by continuity, there must exist an ϵ such that $\forall x \in (\underline{x}, \underline{x} + \epsilon) : h^{(\hat{n})}(x) < 0$. But then by a Taylor expansion up to degree \hat{n} we have that $\forall x \in (\underline{x}, \underline{x} + \epsilon) :$

$$\begin{aligned} H(x) &= \sum_{n=0}^{\hat{n}} \frac{(x - \underline{x})^n}{n!} H^{(n)}(\underline{x}) + \frac{(x - \underline{x})^{\hat{n}+1}}{(\hat{n} + 1)!} H^{(\hat{n}+1)}(y) \quad \text{for some } y \in (\underline{x}, x) \subset (\underline{x}, \underline{x} + \epsilon) \\ &= \frac{(x - \underline{x})^{\hat{n}+1}}{(\hat{n} + 1)!} h^{(\hat{n})}(y) \quad \text{for some } y \in (\underline{x}, \underline{x} + \epsilon) \\ &< 0 \end{aligned}$$

which contradicts the assumption. This concludes the second part of the proof since an argument as the above implies that $h(x) > 0$. For the first part, note that $H(x) = 0$ on $[0, \underline{x})$ and thus $h(x) = 0$ on that same interval. \blacksquare

Lemma 3 *If F MPSPD $_m$ G and both are analytic then $\exists \underline{x} < \bar{x}_m$ such that $\forall x \geq \underline{x}, (-1)^m(F - G) \geq 0$ and for x with $\underline{x} \leq x < \bar{x}_m, (-1)^m[F(x) - G(x)] > 0$.*

Proof : We first prove a preliminary result. Fix some k with $1 \leq k < m$. We will first show that if F MPSPD $_m$ G and $\exists x_{m-k} < 1$ s.t. $\forall x \geq x_{m-k} : (-1)^{m-k}[F_{m-k}(x) - G_{m-k}(x)] \geq 0$ (resp. ≤ 0) with strict inequality for some x then $\exists x_{m-k-1} < 1$ s.t. $\forall x \geq x_{m-k-1} : (-1)^{m-k-1}[F_{m-k-1}(x) - G_{m-k-1}(x)] \geq 0$ (resp. ≤ 0) with strict inequality for some x .

To that purpose, combine the premise with the fact that $\int_0^1 F_{m-k-1}(t) - G_{m-k-1}(t) dt = 0$ (since F MPSPD $_m$ G) to conclude that $\exists x_{m-k} < 1$ s.t. $\forall x \geq x_{m-k} : (-1)^{m-k} \int_x^1 F_{m-k-1}(t) - G_{m-k-1}(t) dt \leq 0$ with strict inequality for some x . Let now $\tilde{x}_{m-k} = \sup(x \leq 1 : (-1)^{m-k} \int_x^1 F_{m-k-1}(t) - G_{m-k-1}(t) dt < 0)$ and note that $\forall x \geq \tilde{x}_{m-k}, (-1)^{m-k} \int_x^1 F_{m-k-1} - G_{m-k-1} = 0$ which implies that $\forall x \geq \tilde{x}_{m-k}, F_{m-k-1}(x) = G_{m-k-1}(x)$. Then, by lemma 2, $\exists \hat{x}_{m-k} < \tilde{x}_{m-k}$ s.t. $\forall x$ with $\hat{x}_{m-k} < x < \tilde{x}_{m-k}, (-1)^{m-k}[F_{m-k-1}(x) - G_{m-k-1}(x)] \leq 0$ with strict inequality for some x . Since $G_{m-k-1}(x) = F_{m-k-1}(x) \forall x \geq \tilde{x}_{m-k}$, we can conclude that $\exists \hat{x}_{m-k} < \tilde{x}_{m-k}$ s.t. $(-1)^{m-k}[F_{m-k-1}(x) - G_{m-k-1}(x)] \leq 0$ with strict inequality for some x , which is what we had to prove for the preliminary result (setting $x_{m-k-1} = \hat{x}_{m-k}$).

Note now that F MPSPD $_m$ G implies that $F_{m-1}(x) - G_{m-1}(x) \leq 0$ with strict inequality for some x . If m is even, this is the same as $(-1)^{m-1}[F_{m-1}(x) - G_{m-1}(x)] \geq 0$ for all x with strict inequality for some. Repeated application of the preliminary result then implies that $\exists \underline{x}_1 < 1$ such that for $x \geq \underline{x}_1 : (-1)^m[F(x) - G(x)] \geq 0$ with strict inequality for some. Using in the very last step (of this repeated application) the second part of lemma 2 shows that $\exists \underline{x}$ such that $\forall x \in (\underline{x}, \bar{x}_m), (-1)^m(F(x) - G(x)) > 0$. If m is odd an analogous argument gives the same result. This implies the proposition. \blacksquare

Lemma 4 *Let F and G be analytic. If $\exists \underline{x} < \bar{x}$ s.t. $\forall x \geq \underline{x} : F(x) - G(x) \geq 0$ (resp. ≤ 0) and $\forall x$ with $\underline{x} \leq x < \bar{x} : F(x) - G(x) > 0$ (resp. < 0), then $\exists N$ such that $\forall n \geq N, E_{F^n}[z(x)] - E_{G^n}[z(x)] < 0$ (resp. > 0) for any $z(x)$ that is continuously differentiable and strictly increasing on $[0, 1]$. In particular, $E_{F^n}[x] - E_{G^n}[x] < 0$.*

Proof : Note first that

$$\begin{aligned}
E_{G^n}[z(x)] - E_{F^n}[z(x)] &= \int_0^1 z(x) d(G^n(x) - F^n(x)) \\
&= [z(x)(G^n(x) - F^n(x))]_0^1 - \int_0^1 z'(x)[G^n(x) - F^n(x)]dx \\
&= \int_0^1 z'(x)[F^n(x) - G^n(x)]dx
\end{aligned}$$

Let $\underline{x}_1 = \inf(x : \forall y, x \leq y < \bar{x}, F(y) - G(y) > 0)$. It is sufficient to show that, $\exists N$ such that $\forall n \geq N$,

$$\int_0^1 z'(x)(F(x) - G(x))F^k(x)G^l(x)dx \geq 0 \quad \text{for } k + l = n - 1, k, l \geq 0$$

Note also that, given the premise of the proposition, $\exists \hat{x}$ with $\underline{x}_1 < \hat{x} < \bar{x}$, $F(\hat{x}) - G(\hat{x}) > 0$ and $F(\hat{x}) < F(\bar{x})$. Define now $\gamma = \frac{F(\hat{x}) + F(\bar{x})}{2F(\hat{x})} > 1$ so that $\gamma F(\hat{x}) = \frac{F(\hat{x}) + F(\bar{x})}{2} < F(\bar{x})$.

Let further $\hat{\hat{x}}$ be defined $\hat{\hat{x}} = \min(y \geq \hat{x} \mid F(y) \geq \gamma F(\hat{x}), G(y) \geq \gamma G(\hat{x}))$. It follows that $F(\hat{\hat{x}})^k G(\hat{\hat{x}})^l \geq \gamma^{k+l} F(\hat{x})^k G(\hat{x})^l$. Note also that $\hat{\hat{x}} < \bar{x}$.

Finally note that the fact that z is continuously differentiable and strictly increasing on $[0, 1]$ implies that $\exists \underline{z}, \bar{z}$ with $0 < \underline{z} \leq \bar{z} < \infty$ such that $\underline{z} \leq z'(x) \leq \bar{z} \forall x \in [0, 1]$.

All this implies that

$$\begin{aligned}
&\int_0^1 z'(x)[F(x) - G(x)]F^k(x)G^l(x)dx \\
&= \int_0^{\hat{\hat{x}}} z'(x)[F(x) - G(x)]F^k(x)G^l(x)dx + \int_{\hat{\hat{x}}}^{\hat{x}} z'(x)[F(x) - G(x)]F^k(x)G^l(x)dx \\
&\quad + \int_{\hat{\hat{x}}}^1 z'(x)[F(x) - G(x)]F^k(x)G^l(x)dx \\
&\geq -1 \int_0^{\hat{\hat{x}}} z'(x)F^k(x)G^l(x)dx + 0 + \int_{\hat{\hat{x}}}^1 z'(x)[F(x) - G(x)]F^k(x)G^l(x)dx \\
&\geq -F^k(\hat{\hat{x}})G^l(\hat{\hat{x}})\bar{z} \int_0^{\hat{\hat{x}}} dx + F^k(\hat{\hat{x}})G^l(\hat{\hat{x}})\underline{z} \int_{\hat{\hat{x}}}^1 [F(x) - G(x)]dx \\
&\geq -F^k(\hat{\hat{x}})G^l(\hat{\hat{x}})\hat{\hat{x}}\bar{z} + \gamma^{k+l}F^k(\hat{x})G^l(\hat{x})\underline{z} \int_{\hat{\hat{x}}}^1 [F(x) - G(x)]dx \\
&= F^k(\hat{\hat{x}})G^l(\hat{\hat{x}}) \left[\gamma^{k+l}\underline{z} \int_{\hat{\hat{x}}}^1 [F(x) - G(x)]dx - \hat{\hat{x}}\bar{z} \right] \\
&> 0 \quad \text{for } k + l \text{ large enough}
\end{aligned}$$

This proves the lemma. ■

Proof of Proposition 3: For the first part of the proposition, let μ denote the probability measure associated with distribution function F . By the definition of the support, $\exists \epsilon > 0, \exists T \subset S$ such that $\mu(T) > \epsilon$ and $\forall (p_n^1, p_n^2) \in T$ $p_n^1 > \bar{p}^2$. For such a choice of ϵ , $P[p_n^1 > \bar{p}^2] \geq \epsilon$, so that $P[p_{Y_1}^1 > \bar{p}^2] = 1 - (1 - \epsilon)^N \rightarrow 1$ as $N \rightarrow \infty$.

The above implies that for the following parts of the proposition, the result holds for $\bar{p}^1 > \bar{p}^2$, so that we can limit our attention in what follows to the situation that $\bar{p}^1 = \bar{p}^2 = \bar{p}$.

Consider now the second part of the proposition. As always, $p_{Y_1}^1 \xrightarrow{\text{a.s.}} \bar{p}$ as $N \rightarrow \infty$. Moreover, since the distribution of p_n^2 has by assumption a density function and thus no atoms, $P[p_n^2 < p] \rightarrow 1$ as $p \rightarrow \bar{p}$. But this implies that $P[p_{Y_1}^1 > p_{Y_1}^2] \xrightarrow{\text{a.s.}} 1$ as $N \rightarrow \infty$.

For the last part of the proposition, it is still true that $p_{Y_1}^1 \xrightarrow{\text{a.s.}} \bar{p}$ as $N \rightarrow \infty$. If $\max_{(\bar{p}, p_n^2) \in S} p_n^2 < \bar{p}$ then the result follows directly by an argument analogous to that for the first point. Assume therefore that $(\bar{p}, \bar{p}) \in S$. Let now $\bar{f} = f(\bar{p}, \bar{p})$. By the continuity of f (and using d for the Euclidean distance), $\forall \epsilon \exists \delta$ s.t. $\forall (p_n^1, p_n^2) \in S : d((p_n^1, p_n^2), (\bar{p}, \bar{p})) < \delta \Rightarrow |f(p_n^1, p_n^2) - f(\bar{p}, \bar{p})| < \epsilon$. Since $\bar{f} > 0$, we can set $\epsilon = \frac{\bar{f}}{2}$. It then follows that $\exists \delta > 0$ such that $\forall (p_n^1, p_n^2) \in S : d((p_n^1, p_n^2), (\bar{p}, \bar{p})) < \delta \Rightarrow \frac{\bar{f}}{2} \leq f(p_n^1, p_n^2) \leq \frac{3\bar{f}}{2}$. Let now $\gamma = \frac{\sqrt{2}\delta}{2}$. For any \tilde{p} with $\bar{p} - \gamma < \tilde{p} < \bar{p}$

$$P[p_n^1 > p_n^2 \mid p_n^1 > \tilde{p}] = \frac{\int_{\tilde{p}}^{\bar{p}} \int_0^{\tilde{p}} dF}{\int_{\tilde{p}}^{\bar{p}} \int_0^{\bar{p}} dF} = 1 - \frac{\int_{\tilde{p}}^{\bar{p}} \int_{p_n^1}^{\bar{p}} dF}{\int_{\tilde{p}}^{\bar{p}} \int_0^{\bar{p}} dF}$$

Let now $\eta = \bar{p} - \tilde{p}$, then $\int_{\tilde{p}}^{\bar{p}} \int_{p_n^1}^{\bar{p}} dF \leq \int_{\tilde{p}}^{\bar{p}} \int_{\tilde{p}}^{\bar{p}} dF \leq \frac{\eta^2 3\bar{f}}{2}$ and $\int_{\tilde{p}}^{\bar{p}} \int_0^{\bar{p}} dF \geq \int_{\tilde{p}}^{\bar{p}} \int_{\bar{p}-\gamma}^{\bar{p}} dF \geq \frac{\eta \gamma \bar{f}}{2}$ so that $\frac{\int_{\tilde{p}}^{\bar{p}} \int_{p_n^1}^{\bar{p}} dF}{\int_{\tilde{p}}^{\bar{p}} \int_0^{\bar{p}} dF} \leq \frac{\eta^2 3\bar{f}/2}{\eta \gamma \bar{f}/2} = \frac{3\eta}{\gamma}$ which goes to zero as $\eta \rightarrow 0$. But this implies that $P[p_n^1 > p_n^2 \mid p_n^1 > \tilde{p}] \rightarrow 1$ as $\tilde{p} \rightarrow \bar{p}$. From which the last part of the proposition follows given that $P[p_{Y_1}^1 > \bar{p} - \epsilon] \rightarrow 1$ as $N \rightarrow \infty$. ■

C.2 Skill or luck

In what follows, use ν_i for the belief of agent i about any event (except for the prior beliefs about the actions that were defined earlier).

Proof of Proposition 8: The first part of the proposition is a special case of the second.

To prove that second part, we have to show that $E[P_{i,A|S_i} - P_{j,A|S_i}]$ is larger under G than under F .

The n 'th derivative of $P_{j,A|S_i} = \frac{p_{Y_i}^j q}{p_{Y_i}^j q + p(1-q)}$ satisfies. $(-1)^{n-1} \frac{d^n P_{j,A|S_i}}{dp_{Y_i}^j} > 0$. Given that $F \text{ MPSPD}_m G$ it follows from proposition 2 in appendix A that $E[P_{j,A|S_i}]$ is *smaller* under G so that it suffices to show that $E[P_{i,A|S_i}]$ is larger. To that purpose, note that lemma 3, lemma 4, and the fact that $P_{i,A|S_i}$ is continuously differentiable and strictly increasing in $p_{Y_i}^i$ imply that $\exists \tilde{N}$ such that $\forall n \geq \tilde{N} : E_{F^n}[P_{i,A|S_i}] < E_{G^n}[P_{i,A|S_i}]$. This implies the second part of the proposition. ■

Proof of Proposition 9: Consider the first part of the proposition and use a instead of p . $\log P(a, l)$ being submodular in a and l implies that $\frac{P_a(a, l)}{P(a, l)}$ decreases in l . Thus for any $k < l$ $\frac{P_a(a, l)}{P(a, l)} < \frac{P_a(a, k)}{P(a, k)}$ or $P(a, k) \frac{P_a(a, l)}{P(a, l)} < P_a(a, k)$ or $\int_0^l P(a, k) \frac{P_a(a, l)}{P(a, l)} dk < \int_0^l P_a(a, k) dk$ or $P_a(a, l) \int_0^l P(a, k) dk < P(a, l) \int_0^l P_a(a, k) dk$. In that case

$$\begin{aligned} \frac{d^2 \log \int_0^l P(a, k) dk}{dadl} &= \frac{d}{da} \left[\frac{d}{dl} \log \int_0^l P(a, k) dk \right] = \frac{d}{da} \frac{P(a, l)}{\int_0^l P(a, k) dk} \\ &= \frac{P_a(a, l) \int_0^l P(a, k) dk - P(a, l) \int_0^l P_a(a, k) dk}{\left[\int_0^l P(a, k) dk \right]^2} < 0 \end{aligned}$$

This implies that $\frac{\int_0^l P_a(a, k) dk}{\int_0^l P(a, k) dk}$ decreases in l so that for $l < 1$

$$\frac{\int_0^l P_a(a, k) dk}{\int_0^l P(a, k) dk} > \frac{\int_0^1 P_a(a, k) dk}{\int_0^1 P(a, k) dk}$$

But then

$$\begin{aligned} \frac{d}{da} P[l \leq \tilde{l} \mid S] &= \frac{d}{da} \frac{\int_0^{\tilde{l}} P(a, k) dk}{\int_0^1 P(a, k) dk} \\ &= \frac{\int_0^{\tilde{l}} P_a(a, k) dk \int_0^1 P(a, k) dk - \int_0^{\tilde{l}} P(a, k) dk \int_0^1 P_a(a, k) dk}{\left[\int_0^1 P(a, k) dk \right]^2} > 0 \end{aligned}$$

so that a decrease in a causes indeed a FOSD shift in the l assessment.

The case with two possible values is similar but follows much more easily

For the second part of the proposition, note that

$$P[l \geq \hat{l} \mid S] = \frac{\int_{\hat{l}}^1 P(a, k) dk}{\int_0^1 P(a, k) dk}$$

The numerator does not change in a by definition of \hat{l} , while the denominator increases in a . It follows that the overall expression decreases in a . \blacksquare

Proof of Corollary 2: For the first part, note that $\frac{\partial^2 \log P(p, l)}{\partial p \partial l} = \frac{\partial}{\partial l} \frac{P_p(p, l)}{P(p, l)} = \frac{P_{p, l}(p, l) P(p, l) - P_p(p, l) P_l(p, l)}{[P(p, l)]^2} < \frac{P_{p, l}(p, l) P(p, l)}{[P(p, l)]^2}$, from which the result thus follows. For the second specification, note that $\frac{d \log(f(p) \cdot g(l))}{dp} = \frac{f'(p) \cdot g(l)}{f(p) \cdot g(l)} = \frac{f'(p)}{f(p)}$ so that the cross partial equals zero. For the third specification, $\frac{d \log(f(p) + g(l))}{dp} = \frac{f'(p)}{f(p) + g(l)}$ so that the cross partial is strictly negative. \blacksquare

C.2.1 An illusion of control

We defined $q_n^i \sim G$ to be the probability that luck does not intervene and $p_n^i \sim F$ to be the probability that the action is a success.

Proof of Proposition 10: Define the new random variable $\pi_n^i = p + q_n^i(p_n^i - p) \in [0, 1]$, which is the payoff that i expects from choosing action n , and let its distribution function be H , with $\max \text{supp } H = \bar{\pi} = \bar{p} + \bar{q}(\bar{p} - p)$.

Agent i chooses the action with the highest π_n^i . So $P[\pi_{Y_i}^i \leq x] = P[\pi_1^i \leq x, \dots, \pi_N^i \leq x] = H(x)^N$ which converges to zero for each $x < \bar{\pi}$ (since $H(x) < 1$ for $x < \bar{\pi}$). It follows that in the limit as $N \rightarrow \infty$, $\pi_{Y_i}^i \xrightarrow{\text{a.s.}} \bar{\pi}$. But this implies that $p_{Y_i}^i \xrightarrow{\text{a.s.}} \bar{p}$ and $q_{Y_i}^i \xrightarrow{\text{a.s.}} \bar{q}$. It follows that $P[q_{Y_i}^i > q_{Y_i}^j] = G(q_{Y_i}^i) \xrightarrow{\text{a.s.}} 1$ which proves the first part of the proposition.

As for the second part, we know that

$$P_{i, A|S_i} = \frac{p_{Y_i}^i q_{Y_i}^i}{p_{Y_i}^i q_{Y_i}^i + p(1 - q_{Y_i}^i)} \xrightarrow{\text{a.s.}} \frac{1 \cdot 1}{1 \cdot 1 + p \cdot 0} = 1$$

Furthermore

$$P_{i, L|F_i} = \frac{(1 - p)(1 - q_{Y_i}^i)}{(1 - p)(1 - q_{Y_i}^i) + (1 - p_{Y_i}^i)q_{Y_i}^i} = \frac{1 - p}{1 - p + \frac{q_{Y_i}^i}{1 - q_{Y_i}^i}(1 - p_{Y_i}^i)}$$

the limit of which depends indeed on the relative speed of convergence of $q_{Y_i}^i$ and $p_{Y_i}^i$. \blacksquare

C.3 Praise yourself, blame the other

Proof of Proposition 11: Let again ν_i denote i 's posterior beliefs and $\nu_{0,i}$ i 's prior beliefs, then

$$\begin{aligned}\nu_1[X = 1 \mid S] &= \frac{\nu_{0,1}[S \mid X = 1]\nu_{0,1}(X = 1)}{\nu_{0,1}[S \mid X = 1]\nu_{0,1}(X = 1) + \nu_{0,1}[S \mid X = 2]\nu_{0,1}(X = 2)} \\ &= \frac{p_{Y_1}^1}{p_{Y_1}^1 + p_{Y_1}^2}\end{aligned}$$

When $N = 1$, the distributions of $p_{Y_1}^1$, $p_{Y_1}^2$, $p_{Y_2}^1$, and $p_{Y_2}^2$ are identical. This implies $\tau_1^S = 1/2$ since $E[\nu_1[X = 1 \mid S]] = \frac{E[\nu_1[X=1|S]] + E[\nu_2[X=2|S]]}{2}$ by symmetry and that equals $\frac{1}{2} \int \frac{p_{Y_1}^1}{p_{Y_1}^1 + p_{Y_1}^2} + \frac{1}{2} \int \frac{p_{Y_2}^2}{p_{Y_2}^2 + p_{Y_2}^1} = \frac{1}{2} \int \frac{p_{Y_1}^1}{p_{Y_1}^1 + p_{Y_1}^2} + \frac{1}{2} \int \frac{p_{Y_1}^2}{p_{Y_1}^2 + p_{Y_1}^1} = \frac{1}{2}$.

Note further that $\frac{p_{Y_1}^1}{p_{Y_1}^1 + p_{Y_1}^2}$ strictly increases in $p_{Y_1}^1$ and F^{N+1} strictly FOSD $F^N \forall N > 1$. It follows that for $N > 1$, $E[\tau_1^S] > 1/2$ and increases in N .

Finally, as $N \rightarrow \infty$, $\nu_1[X = 1 \mid S] \xrightarrow{\text{a.s.}} \frac{\bar{p}}{\bar{p} + p_{Y_1}^2} > 1/2$ a.s..

This implies the ‘success’ parts of the proposition. The ‘failure’ parts are analogous. ■

C.4 Everyone is better than average?

Proof of Proposition 12: Consider an individual with belief $\alpha_y \in [0, 1]$. Note that this person is almost surely the only one with this particular belief. This individual’s problem in choosing his attention levels is $\max_{a_r, a_s \geq 0} \alpha_y r(a_r) + (1 - \alpha_y)s(a_s)$ (a.s.) s.t. $a_r + a_s \leq \bar{a}$.

This problem has a unique solution, denoted $(a_{r,y}, \bar{a} - a_{r,y})$ with $a_{r,y}$ strictly increasing in α_y , by the implicit function theorem. It follows that if $\alpha_y \neq \alpha_x$ then $a_{r,y} \neq a_{r,x}$, so that by the definition of $(a_{r,x}, a_{s,x})$ and the uniqueness of the maximum $\alpha_x r(a_{r,x}) + (1 - \alpha_x)s(a_{s,x}) > \alpha_x r(a_{r,y}) + (1 - \alpha_x)s(a_{s,y})$. So this person with belief α_y thinks his action is strictly better than that of any other person with some belief α_x . ■

References

- ADAMS, J. (1965): "Inequity in Social Exchanges," in *Advances in Experimental Social Psychology*, ed. by L. Berkowitz. Academic Press, New York.
- ALLISON, S., D. MESSICK, AND G. GOETHALS (1989): "On Being Better but not Smarter than Others: The Mohammed Ali Effect," *Social Cognition*, 7, 275– 296.
- AUMANN, R. J. (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1– 18.
- AUMANN, R. J. (1998): "Common Priors : A Reply to Gul," *Econometrica*, 66(4), 929– 938.
- AUMANN, R. J., AND A. BRANDENBURGER (1995): "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63, 1161– 1180.
- BABCOCK, L., AND G. LOEWENSTEIN (1997): "Explaining Bargaining Impasse: The Role of Self-Serving Biases," *Journal of Economic Perspectives*, 11(1), 109– 126.
- BABCOCK, L., G. LOEWENSTEIN, AND S. ISSACHAROFF (1995): "Biased Judgements of Fairness in Bargaining," *American Economic Review*, 85(5), 1337– 1343.
- BAKER, G., M. JENSEN, AND K. MURPHY (1988): "Compensation and Incentives: Practice versus Theory," *Journal of Finance*, 43(3), 593– 616.
- BAUMEISTER, R. F. (1998): "The Self," in *The Handbook of Social Psychology*, ed. by D. T. Gilbert, S. T. Fiske, and G. Lindzey. McGraw-Hill, New York.
- BAZERMAN, M. (1998): *Judgment in Managerial Decision Making*. John Wiley & Sons, New York.
- BENABOU, R., AND J. TIROLE (1999): "Self-Confidence: Intrapersonal Strategies," Working Paper, SSRN.
- BETTMAN, J., AND B. WEITZ (1983): "Attributions in the Board Room: Causal Reasoning in Corporate Annual Reports," *Administrative Science Quarterly*, 28, 165– 183.
- BLAINE, B., AND J. CROCKER (1993): "Self-esteem and Self-serving Biases in Reactions to Positive and Negative Events: An Integrative Review," in *Self-esteem: The Puzzle of Low Self-regard*, ed. by R. Baumeister. Plenum, New York.
- BROWN, K. C. (1974): "A Note on the Apparent Bias of Net Revenue Estimates for Capital Investment Projects," *Journal of Finance*, 29, 1215– 1216.
- CAPEN, E., R. CLAPP, AND W. CAMPBELL (1971): "Competitive Bidding in High-Risk Situations," *Journal of Petroleum Technology*, 23, 641– 653.
- COMPTE, O. (2002): "The Winner's Curse with Independent Private Values," Working Paper CERAS-ENPC.
- COMPTE, O., AND A. POSTLEWAITE (2001): "Confidence Enhanced Performance," Working paper.
- CURREN, M., V. FOLKES, AND J. STECKEL (1992): "Explanation for Successful and Unsuccessful Marketing Decisions: The Decision Maker's Perspective," *Journal of Marketing*, 56, 18– 31.
- DUVAL, T. S., AND P. J. SILVIA (2002): "Self-Awareness, Probability of Improvement, and the Self-serving Bias," *Journal of Personality and Social Psychology*, 82, 49– 61.
- ELLISON, G. (2000): "Evolving Standards for Academic Publishing: A q-r Theory," Working paper.
- FISKE, S., AND S. TAYLOR (1991): *Social Cognition*. McGraw-Hill, New York.
- GOODMAN, P., AND A. FRIEDMAN (1971): "An Examination of Adams' Theory of Inequity," *Administrative Science Quarterly*, 16, 271– 288.
- GUL, F. (1998): "A Comment on Aumann's Bayesian View," *Econometrica*, 66(4), 923– 927.

- HARRISON, J., AND J. MARCH (1984): "Decision Making and Postdecision Surprises," *Administrative Science Quarterly*, 29, 26– 42.
- HARRISON, M., AND D. M. KREPS (1978): "Speculative Investor Behavior in a Stock Market with Heterogenous Expectations," *Quarterly Journal of Economics*, 92(2), 323– 336.
- HARSANYI, J. C. (1967): "Games with Incomplete Information Played by 'Bayesian' Players, Parts I, II and III," *Management Science*, 14, 159– 182.
- KAHNEMAN, D., AND A. TVERSKY (1973): "On the Psychology of Prediction," *Psychological Review*, 80, 237– 251.
- KAPLAN, R. (1986): "Must CIM Be Justified by Faith Alone?," *Harvard Business Review*, (2), 87– 93.
- LANGER, E. J. (1975): "The Illusion of Control," *Journal of Personality and Social Psychology*, 32, 311– 328.
- LARWOOD, L., AND W. WHITTAKER (1977): "Managerial Myopia: Self-Serving Biases in Organizational Planning," *Journal of Applied Psychology*, 62(2), 194– 198.
- LAZEAR, E. (2001): "The Peter Principle: Promotions and Declining Productivity," NBER Working Paper W8094.
- LORD, C., L. ROSS, AND M. LEPPER (1979): "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence," *Journal of Personality and Social Psychology*, 37, 2098– 2109.
- MEYER, H. (1975): "The Pay-for-Performance Dilemma," *Organizational Dynamics*, 3, 39– 50.
- MILES, E., J. HATFIELD, AND R. HUSEMAN (1989): "The Equity Sensitivity Construct: Potential Implications for Worker Performance," *Journal of Management*, 15(4), 581– 588.
- MILLER, D., AND M. ROSS (1975): "Self-serving Biases in the Attribution of Causality: Fact or Fiction?," *Psychological Bulletin*, 82, 213– 225.
- MILLER, D. T. (1976): "Ego Involvement and Attributions for Success and Failure," *Journal of Personality and Social Psychology*, 34, 901– 906.
- MORRIS, S. (1995): "The Common Prior Assumption in Economic Theory," *Economics and Philosophy*, 11, 227– 253.
- NISBETT, R., AND L. ROSS (1980): *Human Inference : Strategies and Shortcomings of Social Judgment*. Prentice Hall, Englewood Cliffs, New Jersey.
- PETER, L., AND R. HULL (1969): *The Peter Principle*. W. Morrow, New York.
- POTERBA, J. M., AND L. H. SUMMERS (1995): "A CEO Survey of U.S. Companies' Time Horizons and Hurdle Rates," *Sloan Management Review*, pp. 43– 53.
- PRIEST, G., AND B. KLEIN (1984): "The Selection of Disputes for Litigation," *Journal of Legal Studies*, 13, 1– 55.
- PYSZCZYNSKI, T., AND J. GREENBERG (1987): "Towards an Integration of Cognitive and Motivational Perspectives on Social Inference: A Biased Hypothesis-testing Model," in *Advances in Experimental Social Psychology* (v20), ed. by L. Berkowitz. Academic Press, San Diego, CA.
- RABIN, M., AND J. SCHRAG (1999): "First Impressions Matter: A Model of Confirmation Bias," *Quarterly Journal of Economics*, 114(1), 37– 82.
- ROSS, M., AND F. SICOLY (1979): "Egocentric Biases in Availability and Attribution," *Journal of Personality and Social Psychology*, 37, 322– 336.
- SEDIKIDES, C., W. K. CAMPBELL, AND G. D. REEDER (1998): "The Self-serving Bias in Relational Context," *Journal of Personality and Social Psychology*, 74(2), 378– 386.

- STREUFERT, S., AND S. C. STREUFERT (1969): "Effects of Conceptual Structure, Failure, and Success on Attribution of Causality and Interpersonal Attitudes," *Journal of Personality and Social Psychology*, 11, 138– 147.
- SVENSON, O. (1981): "Are we all less risky and more skillful than our fellow drivers?," *Acta Psychologica*, 47, 143– 148.
- TAYLOR, S. E., AND J. D. BROWN (1988): "Illusion and Well-being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin*, 103(2), 193– 210.
- VAN DEN STEEN, E. J. (2001): "Essays on the Managerial Implications of Differing Priors," PhD Dissertation, Stanford University.
- VAN DEN STEEN, E. J. (2002): "Overconfidence by Rational Agents," Working Paper.
- WADE-BENZONI, K., L. THOMPSON, AND M. BAZERMAN (2000): "The Malleability of Environmentalism," Working paper, HBS.
- WEINSTEIN, N. D. (1980): "Unrealistic Optimism about Future Life Events," *Journal of Personality and Social Psychology*, 39(5), 806– 820.
- (1982): "Unrealistic Optimism about Susceptibility to Health Problems," *Journal of Behavioral Medicine*, 5(4), 441– 460.
- (1984): "Why It Won't Happen to Me: Perceptions of Risk Factors and Susceptibility," *Health Psychology*, 3(5), 431– 457.
- WILSON, R. (1977): "A Bidding Model of Perfect Competition," *Review of Economic Studies*, 4, 511– 518.
- WOLOSIN, R. J., S. J. SHERMAN, AND A. TILL (1973): "Effects of Cooperation and Competition on Responsibility Attribution after Success and Failure," *Journal of Experimental Social Psychology*, 9, 220– 235.
- YILDIZ, M. (2000): "Sequential Bargaining without a Recognition Device," working paper.
- ZABOJNIK, J. (2000): "A Model of Rational Bias in Self-Assessments," Working paper.
- (2001): "Centralized and Decentralized Decision-Making in Organizations," *Journal of Labor Economics*.
- ZUCKERMAN, M. (1979): "Attribution of success and failure revisited, or: The motivational bias is alive and well in attribution theory," *Journal of Personality*, 47, 245– 287.