Rational Overoptimism (and Other Biases)

By ERIC VAN DEN STEEN*

Rational agents with differing priors tend to be overoptimistic about their chances of success. In particular, an agent who tries to choose the action that is most likely to succeed, is more likely to choose an action of which he overestimated, rather than underestimated, the likelihood of success. After studying the comparative statics of this mechanism, I show that it also causes agents to attribute failure to exogenous factors but success to their own choice of action, to disproportionately believe that they will outperform others, to overestimate the precision of their estimates, and to overestimate their control over the outcome. (*JEL* A12, B49, C70, D81, D84)

Human inference and estimation is subject to systematic biases. The evidence shows, for example, that people are overoptimistic about future life events (Neil D. Weinstein, 1980, 1982, 1984, and the reviews therein), that they tend to attribute success to their own actions but failure to external factors (Miron Zuckerman, 1979; Susan Fiske and Shelley Taylor, 1991; Roy F. Baumeister, 1998; Thomas S. Duval and Paul J. Silvia, 2002), that they overestimate their contribution in joint projects (Michael Ross and Fiore Sicoly, 1979), that they overestimate the precision of their estimates (Stuart Oskamp, 1965), and that they also overestimate the degree to which they have control over an outcome (Ellen J. Langer, 1975).

The economics and psychology literature has suggested various explanations for these phenomena, which I discuss in more detail below. The purpose of this paper is to propose an alternative mechanism that is structurally different from the explanations forwarded by the existing literature.

The basic idea is as follows. If agents sometimes overestimate and sometimes underestimate the probability of success of an action (relative to other agents) and try to select the action with the highest probability of success, then they are more likely to select actions of which they overestimated the probability of success. They will therefore tend to be overoptimistic (relative to these other agents) about the likelihood of success of the actions they undertake. This mechanism is similar to the winner's curse: in both cases, random variation plus systematic choice lead to a systematic bias.

The purpose of this paper is to formalize this "choice-driven overoptimism" mechanism, derive comparative statics, and show that it can lead to the other biases mentioned above. To that purpose, I build a simple model in which an agent has to choose an action, but agents may openly disagree about the likelihood of success of alternative actions. In particular, I assume that the agents' beliefs are independently and identically distributed (i.i.d.). The disagreement combined with optimizing behavior will lead the agent to be positively biased about the consequences of his own actions, relative to others.

The bias increases in the number of distinct alternatives, keeping the dispersion of priors constant while still assuming that the agents' beliefs are independent. The bias also increases in a mean-preserving spread of the distribution of prior beliefs, but it tends to disappear with sufficient experience with the particular choice problem. Finally, the bias will increase in the importance of success if there is complementary

* Sloan School of Management, MIT, Cambridge, MA 02142 (e-mail: evds@mit.edu). John Roberts' suggestions and support made an enormous difference to me. Bengt Holmstrom's remarks also had a big impact on this and other work. I got extensive suggestions from Mark Satterthwaite, Ján Zábojník, and the anonymous referees, and very useful comments from Dan Ariely, Bill Barnett, Max Bazerman, Bob Gibbons, Thomas Hellmann, Ed Lazear, Jon Levin, Sendhil Mullainathan, Paul Oyer, Drazen Prelec, Emily Pronin, Garth Saloner, Antoinette Schoar, Joel Sobel, 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), Stanford Graduate School of Business, and MIT-Sloan are gratefully acknowledged.

effort by the agent. As I argue in Section III, all but the last of these comparative statics empirically distinguish this theory from others. At this point, however, the existing evidence does not allow us to make that distinction.

I then show that this same mechanism can also cause the following biases.

- More than 50 percent of a population believe that they will do better than the median.
- Agents tend to attribute success to their own choice of action and failure to exogenous factors. In a cooperative venture, they tend to attribute success to their own choice of action and failure to the action choice of their associate.
- Agents tend to overestimate the precision of their estimates.
- Agents tend to overestimate the control they have over the outcome.

Literature.—There is an extensive psychology and a rapidly growing economic literature on the topic of judgment biases. The psychology literature provides essentially two types of explanations for these phenomena. The first type are motivational theories, i.e., theories that posit that the biases exist since they (subconsciously) benefit the agent through, for example, self-esteem maintenance. The general idea is that agents hold the beliefs that make them most happy or most successful (Zuckerman, 1979; Baumeister, 1998). Hans K. Hvide (2002) and Markus K. Brunnermeier and Jonathan A. Parker (2003) assume that such beliefs are the outcome of an intentional choice process and develop economic models to study the optimal choice of beliefs, in which the agents trade off current happiness against the cost of future incorrect decisions.

Some authors in the psychology literature have challenged the motivational view by forwarding cognitive explanations, i.e., theories that attribute the biases to information processing effects that are either benefit-neutral or may even hurt the agent's mental health or general well-being. Dale T. Miller and Michael 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. In the end, however, such cognitive biases also seem to be driven by motivation (Zuckerman, 1979). This is made more explicit in theories of "motivated reasoning" (Ziv Kunda, 1990) where people's memory and reasoning is affected by their motivation. A number of economists have followed a similar line by appealing to Bayesian decision makers with some motivated limitation or biases in their information processing capabilities. Matthew Rabin and Joel Schrag (1999), for example, show how a confirmatory bias may lead to overconfidence. Roland Benabou and Jean Tirole (2002) show how self-serving biases can arise if people have (endogenously chosen) imperfect recall. Olivier Compte and Andrew Postlewaite (2003) show that self-serving biases in interpreting or remembering outcomes may be optimal if an agent's performance depends on her self-confidence. Isabelle Brocas and Juan D. Carrillo (2002) and Ján Zábojník (2004) present models in which people can learn about their abilities and, as a consequence, more than half the population end up considering themselves better than the median, although there is no bias on average. Ronit Bodner and Drazen Prelec (2003) study how people might take certain actions to signal to themselves that they possess some desirable but imperfectly observable trait, such as kindness or intelligence, and they conclude that people become overoptimistic if they are not fully aware of their self-signaling incentives.

The key distinction between the model in this paper and these theories is that the current mechanism does not focus on the agent's mental processes but instead on the situation, i.e., on the agent's choice of action. The bias does not originate in the process of deriving beliefs about the consequences of a particular action, but in the choice of which particular belief is relevant, through the agent's choice of action. This will give rise to very different comparative statics and to different remedies.

This choice-driven endogenous overoptimism is structurally similar to the winner's curse (Edward Capen et al., 1971), regression towards the mean (Daniel Kahneman and Amos Tversky, 1973), post-decision surprises (Keith C. Brown, 1974; J. Richard Harrison and James G. March, 1984), and Edward Lazear's (2004) explanation for the Peter Principle (Laurence Peter and Raymond Hull, 1969). Brown (1974), for example, discusses how this kind of mechanism causes investment projects to fall short of their estimated revenue. Luis Santos-Pinto and

Joel Sobel (2003) generalize this model by including investments in skills or technologies and derive some interesting new conclusions. There are also some papers that develop similar arguments in other contexts. In the finance literature, Michael Harrison and David M. Kreps (1978) and Stephen Morris (1996) have appealed to a similar mechanism with heterogenous priors to explain why equity prices might exceed agents' valuations, especially at the time of the equity offering. Olivier Compte (2002) shows that with differing priors, the winner's curse also applies to the independent private values case.

The contribution of this paper is to formalize this overoptimism mechanism in a general form, to demonstrate how it can cause several other well-known judgment biases, and to derive comparative statics that determine when these biases will be most pronounced and allow verification of the model. The analysis also has some very practical implications. It suggests, for example, that many experiments may actually underestimate the importance of the bias. It also suggests some practical measures that can reduce or eliminate the bias.

Section I derives the basic overoptimism result. Section II derives the comparative statics and Section III discusses the role of differing priors and potential tests of the theory. Section IV shows how this mechanism can cause other well-recognized biases. Section V concludes. The Appendix considers generalizations of the theory.

I. Choice-Driven Overoptimism

To formalize the overoptimism result, consider J agents who each choose an action out of a set of N potential actions a_n. The outcome of choosing an action will be either a success or a failure. In choosing the action, the agent tries to maximize the likelihood of having a success. In particular, assume that agents are risk-neutral revenue maximizers and that a success gives a payoff 1, while failure gives payoff 0.

An action's probability of success is uncertain. All agents have subjective beliefs about this probability. These beliefs may differ but are common knowledge, i.e., agents have differing priors.[1] In particular, let F_n^i denote agent i's prior belief distribution regarding the probability of success of action a_n. Assume that F_n^i has full support [0, 1] and mean p_n^i. Note that $p_n^i = \int p \, dF_n^i$ is the overall likelihood of success of action a_n according to agent i. This mean is all that matters to the results. To simplify, I will therefore formulate the *ex ante* distribution of beliefs in terms of the distribution of the p_n^i. In particular, assume that the p_n^i are i.i.d. draws from an atomless distribution G on $[\underline{p}, \bar{p}] \subseteq [0, 1]$, i.e., $p_n^i \sim G[\underline{p}, \bar{p}]$ with $\underline{p} < \bar{p}$. Let $\mathcal{G}$ denote the joint distribution, i.e., $\mathcal{G} = \times_N G$. Note that G is *not* some kind of prior and agents do *not* have any private information. The distribution G just represents the empirical variation in the beliefs of the agents.[2]

The key 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 ρ_i^j denote the likelihood according to agent j that agent i will be successful (where we allow $j = i$).

PROPOSITION 1: *When the number of actions $N > 1$, $\forall j \neq 1$,*

- $E_{\mathcal{G}}[\rho_1^1 - \rho_1^j] > 0$
- $\rho_1^1 - \rho_j^1 \geq 0$, *while* $E_{\mathcal{G}}[\rho_1^1 - \rho_j^1] > 0$.

PROOF:

The distribution of $p_{Y_1}^1$, a first-order statistic, is $G^N(p)$. The distribution of $p_{Y_1}^j$ is $G(p)$, since Y_1 is a randomly selected action from j's perspective. By integration by parts, $E_{\mathcal{G}}[\rho_1^1 - \rho_1^j] = \int_{\underline{p}}^{\bar{p}} G(x) - G^N(x) \, dx$. This equals zero when $N = 1$ and is strictly larger than zero when $N > 1$. The second part of the proposition follows directly from revealed preference and the fact that $E_{\mathcal{G}}[\rho_1^j] = E_{\mathcal{G}}[\rho_j^1]$.

[1] 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. John C. Harsanyi (1968) 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." For a more extensive discussion, see Morris (1995) or Van den Steen (2001).

[2] Whether agents know G is immaterial to the model. Note, though, that belief *realizations* are assumed to be common knowledge.

The first part of the proposition says that the expected overoptimism of one agent according to another is strictly positive, as long as there is more than one action. If agent 1 is the CEO of a company and j is a shareholder, then the CEO will be overoptimistic about the probability of success of his strategy, according to that shareholder. The second part of the proposition says that any agent expects to do better than others. A direct consequence of the second part is that more than 50 percent of the population expect to do better than the median. If driver quality is measured by the probability of having an accident, the set of actions are different driving styles, and "success" is the event of avoiding an accident, then more than 50 percent of the population will believe that they drive better than the median driver.

Note also that, following the proof of the proposition, $E_{\mathcal{G}}[\rho_1^1 - \rho_1^j] = 0$ when the number of actions $N = 1$. The agent will only be overoptimistic when she has a choice of action. This highlights the key role of choice in this model and has three important implications.

1. By eliminating choice, we can eliminate the overoptimism generated by this mechanism. This suggests two ways to reduce the bias: the separation of decision rights and the use of outsiders to estimate the likelihood of success after a project has been selected.
2. Experimental studies that limit the agents' freedom of action, as they often do, will underestimate the magnitude of the bias. This is related to the notion of "weak" versus "strong" environments (Walter Mischel, 1977; Mark Snyder and William J. Ickes, 1985).[3]
3. Whenever agents are overoptimistic but did not have a choice, a different mechanism must be at work.

The Appendix shows that this basic result holds for a wide class of belief distributions and that it can be generalized to overestimation of expected utility for very general utilities, choice spaces, and state spaces.

II. Comparative Statics

This section derives four comparative statics. The bias increases in the number of distinct alternatives and in the variance of the beliefs, and disappears with sufficient experience with the particular decision at hand. It also increases in the importance of the decision when effort is complementary to success.

A. *Number of Distinct Alternatives*

The first result says that both the agent's overoptimism according to others and the degree to which he believes that he will do better than others, increase in the number of distinct alternatives, keeping the dispersion of beliefs for each alternative constant.

PROPOSITION 2A: $\frac{dE_{\mathcal{G}}[\rho_1^1 - \rho_1^j]}{dN} > 0$ *and* $\frac{dE_{\mathcal{G}}[\rho_1^1 - \rho_j^1]}{dN} > 0.$

PROOF:

Since $E_{\mathcal{G}}[\rho_1^1 - \rho_1^j] = \int_{\underline{p}}^{\bar{p}} G(x) - G^N(x)\, dx$ and $\forall x \in (\underline{p}, \bar{p})$, $G^N(x)$ decreases strictly in N, the result is immediate. The second part follows as before.

The assumptions that the realizations are independent and that the dispersion of beliefs over each action is kept constant are important. If we created more replicas of the same actions, and thus of the same beliefs, the overoptimism would not increase. This is important for empirical and experimental tests of the theory.

This comparative static also has practical implications. An agent who can choose his own projects will be more optimistic than one who gets assigned his projects. This may be one reason why entrepreneurs often seem more overoptimistic than regular employees. It also implies that restricting a manager's degrees of freedom may reduce her bias.

For an experimental test of the theory, it is useful to derive a slightly different but closely

[3] A "strong" setting is one that "lead[s] everyone to construe the particular events in the same way, induce uniform expectancies regarding the most appropriate response pattern, [...]" (Mischel, 1977), while "weak" settings are exactly the opposite. Ickes (1982) and Snyder and Ickes (1985) have argued that experiments have traditionally focused on "strong" settings and that shifting the focus to weak settings would lead to more significant results.

related result. Let Y_i again denote the action choice of agent i.

PROPOSITION 2B: $$\frac{dP[\rho_1^1 > \rho_1^j | Y_1 \neq Y_j]}{dN} > 0.$$

PROOF:

Combining $P[\rho_1^1 > \rho_1^j] = \int_0^1 [\int_u^1 dG^N(v)]g(u)\,du = \frac{N}{N+1}$ and $P[\rho_1^1 > \rho_1^j] = P[\rho_1^1 > \rho_1^j | Y_1 = Y_j]P[Y_1 = Y_j] + P[\rho_1^1 > \rho_1^j | Y_1 \neq Y_j]P[Y_1 \neq Y_j]$ gives $P[\rho_1^1 > \rho_1^j | Y_1 \neq Y_j] = \frac{\frac{N}{N+1} - \frac{1}{2N}}{1 - \frac{1}{N}} = 1 - \frac{1}{2(N+1)}$ which increases in N.

This result says that the probability that an agent is considered overoptimistic by some other agent, conditional on the two agents choosing different actions, increases in the number of distinct alternatives. This comparative static is robust to actions having different belief distributions, as long as these distributions are still independent. This makes it very useful for experimental tests. The reason for conditioning on the agents' choosing different actions is to make sure that this comparative static distinguishes this theory from the others.

The following proposition gives some idea where the increase in alternatives can lead. It says that, in the limit, the probability of being considered overoptimistic approaches 1.

PROPOSITION 2C: *For any $j \neq 1$, as $N \to \infty$, $P[\rho_1^1 - \rho_1^j > 0] \to 1$.*

PROOF:

$P[\rho_1^1 > \rho_1^j] = \int_{\underline{p}}^{\bar{p}} 1 - P[\rho_1^1 < x]\,dG(x) = \int_{\underline{p}}^{\bar{p}} (1 - G^N(x))\,dG(x) \xrightarrow{N\to\infty} 1.$

B. *Variance of Belief Distribution*

The overoptimism also increases in a mean-preserving spread of the distribution of the priors.

PROPOSITION 3: *If H is a mean-preserving spread of G, then $E_{\mathcal{H}}[\rho_1^1 - \rho_1^j] > E_{\mathcal{G}}[\rho_1^1 - \rho_1^j]$ and $E_{\mathcal{H}}[\rho_1^1 - \rho_j^1] > E_{\mathcal{G}}[\rho_1^1 - \rho_j^1]$ $\forall N > 1$.*

PROOF:

Since H is a mean-preserving spread of G, $E_{\mathcal{H}}[\rho_1^j] = E_{\mathcal{G}}[\rho_1^j]$ and idem for ρ_j^1. It thus suffices to show that $E_{\mathcal{H}}[\rho_1^1] > E_{\mathcal{G}}[\rho_1^1]$ or $[1 - \int_0^1 H^N] > [1 - \int_0^1 G^N]$. For any k such that $0 \leq k \leq N-1$, integration by parts implies

$$\int_0^1 (G - H)G^{N-k-1}H^k\,dx$$

$$= \left[G^{N-k-1}H^k \int_0^y (G - H)\,dx \right]_0^1$$

$$- \int_0^1 \left\{ \int_0^y (G - H)\,dx \right\} d(G^{N-k-1}H^k)$$

$$= -\int_0^1 \left\{ \int_0^y (G - H)\,dx \right\} d(G^{N-k-1}H^k) > 0$$

by the definition of mean-preserving spread and the fact that $G^{N-k-1}H^k$ is a distribution function. Repeated application implies the result.

In other words, there will be no bias when all agents have a common prior, and the bias will increase as there is more open disagreement. We would therefore conjecture that managerial overoptimism is more prevalent in industries where there is a lot of disagreement about the merits of alternative strategies or alternative technologies. Proposition 4 suggests that this would be more typical for new industries or new technologies.

An earlier working paper, Van den Steen (2002), obtains results on all the higher-order moments of the priors distribution using a more general definition of stochastic dominance. It shows, for example, that the bias will be stronger when the distribution is positively skewed.

C. *Previous Experience*

It also seems intuitive that the bias would disappear as the agents gain more experience with the alternatives, since the disagreement will decrease with experience. To make this more precise, I need to modify the setting to a repeated version of the game in Section I. Consider therefore an infinitely repeated choice problem, with a discount factor $\delta \in (0, 1)$. The prior beliefs of the agents are as specified in Section I. For simplicity, assume that in each period, agent 1 (and agent 1 only) publicly chooses an action a_n out of a set of N actions. At the end of the period, all agents learn the outcome. In this case, the true probabilities of success will play a role. Assume that they are also independent draws from the G distribution. This situation is the well-known multi-armed bandit problem with independent arms and geometric discounting. Let the index t denote the period. Let $(\Omega, \mathcal{F}, P)$ denote the probability space generated by the outcomes. The following proposition says that in the limit the agent's perceived overoptimism about his own action disappears.

PROPOSITION 4: *As $t \to \infty$, $\rho^1_{1,t} \to \rho^j_{1,t}$ with P-probability one.*

PROOF:

Note first that, by Michael Rothschild (1974), the agent will eventually settle on an action with probability 1. Since each prior had full support, merging (David Blackwell and Lester E. Dubins, 1962) implies that both $\rho^1_{1,t} \to \hat{p}$ and $\rho^j_{1,t} \to \hat{p}$ with P-probability 1 so that $\rho^1_{1,t} \to \rho^j_{1,t}$ with P-probability 1.

Since I limited the setting to one agent taking actions, there is some probability that in the limit the agent still thinks his own action is better than what some other agent would undertake. The proposition also does not say that the convergence is monotone over time. In fact, it is possible to find counterexamples to such monotone convergence, although they seem to be fairly extreme cases. Note also that biases need not disappear if all employees have a finite lifetime and are succeeded by new employees who learn imperfectly about the past. While Proposition 4 suggests that managerial overoptimism may be more prevalent in new industries and for new technologies, we have to be careful with this interpretation due to the possible lack of monotone convergence.

D. *Importance of Success in the Presence of Complementary Effort*

Consider now the following extension. Assume that success also depends on the agent's effort. In particular, if the agent spends effort $e \in [0, 1]$ and undertakes action a_n, then he believes that his expected probability of success is ep^i_n. Let the agent's cost of effort $c(e)$ be strictly increasing and strictly convex with $c(0) = 0$ and $c'(0) = 0$. Let his benefit from a success be $\gamma > 0$ and his benefit from a failure be 0. Let ρ^i_j include the effect of effort.

PROPOSITION 5: *The expected biases $E_{\mathcal{G}}[\rho^1_1 - \rho^j_1]$ and $E_{\mathcal{G}}[\rho^1_1 - \rho^1_j]$ increase in γ.*

PROOF:

For the first part, assume that agent 1 considers a_n to be optimal. He then solves $\max_{e \in [0,1]} \gamma e p^i_n - c(e)$ which has a unique solution $\hat{e}$ that increases in γ. The ρ^1_1 and ρ^j_1 are now respectively $\hat{e}p^i_n$ and $\hat{e}p^j_n$, so that the difference increases in γ. The second part is analogous.

This result captures the intuitive notion that on less important things, agents will spend less effort and thus have less reason to become overoptimistic or believe that they will do better than others. It is important to note that the result is sensitive to the specification. The specification above, however, seems to be the most natural. Note also that this comparative static does not depend on the process by which the bias is generated. It can therefore not be used to distinguish this model from others.

III. Discussion

Differing Priors.—A natural question is whether we need differing priors to obtain the mechanism in this paper. 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, to take into account this bias, his estimate will be lower than his signal. As a consequence, it cannot happen, for example, that on average more than 50 percent of the agents expect to do better than the median. The reason why such "de-biasing" does not happen with differing priors is that (by definition) the agent is convinced that his prior is correct and that all other people are wrong.

The agent's signal *is* his belief. He does not see a need to de-bias.

Potential Tests of the Theory.—The existing body of experimental evidence does not allow us to clearly distinguish this theory from the motivational and cognitive theories. On the other hand, however, the theory does make predictions that allow such distinction; in particular the predictions regarding the effect of the number of distinct alternatives, the effect of a mean-preserving spread of the priors distribution, and the effect of experience.

To see that these predictions indeed distinguish this theory from the others, note that the pure motivational theories are forward looking and relate only to what the agent ends up doing. As a consequence, their conclusions are independent of the number of alternatives, the priors of other agents, and any past experience. In the cognitive theories, the number of alternatives and the priors of other agents also have no role. Past experience, however, is a key factor, although none of these theories seems to predict that the bias will decrease with experience. On the contrary, it is precisely the past experiences that create the bias in, for example, Benabou and Tirole (2002).

IV. Relationship to Other Biases

This mechanism can also cause some other well-recognized biases. In analyzing these, I will work with simple variations on the original model. While this approach limits the scope of the results, it exposes most clearly the links between the different models.

A. *Asymmetric Attribution of Causality of Success and Failure*

An important further effect of the mechanism is that it also leads to biased attributions of causality. In particular, from another agent's perspective, the agent will underestimate the role of exogenous factors when he has a success and overestimate their role in case of failure.

To see this more formally, extend the model in Section I to allow the likelihood of success to depend both on the action choice and on an exogenous factor. In particular, let the probability of success be $\alpha p_n + (1 - \alpha)\varepsilon$ where $\varepsilon \in (0, 1)$ is fixed and known, and represents an exogenous factor. As before, p_n is a probability of success that still depends on the chosen action, a_n. The factor α is a random variable that is not observed but is commonly known to have a distribution without atoms, H, on $[\underline{\alpha}, \bar{\alpha}] \subset [0, 1]$. The factor α determines the degree to which the agent influences the likelihood of success via his choice of action. With $\alpha = 0$, the agent cannot influence the outcome at all, while with $\alpha = 1$, the likelihood of success depends completely on the agent's choice of action. As before, agents do not know p_n but form beliefs. They also do not observe α but update their beliefs about α based on the outcome. Let $\hat{H}^i$, with mean $\hat{\alpha}^i$, be the posterior distribution of α according to agent i. If $\hat{\alpha}^1 > \hat{\alpha}^i$ then, according to agent i, agent 1 underestimates the role of the exogenous random factor in the outcome. The following proposition says that, from i's perspective, agent 1 tends to underestimate the role of the exogenous factor when he has a success and overestimate it when he has a failure.

PROPOSITION 6: *Let $N > 1$. After observing a success $E_{\mathcal{G}}[\hat{\alpha}^1 - \hat{\alpha}^i] > 0$, while after observing a failure $E_{\mathcal{G}}[\hat{\alpha}^1 - \hat{\alpha}^i] < 0$.*

PROOF:

In case of a success, $\hat{h}(\alpha|S, p) = \dfrac{P[S|\alpha, p]h(\alpha|p)}{\int_0^1 P[S|\alpha, p]h(\alpha|p)\, d\alpha} = \dfrac{(\varepsilon + \alpha(p - \varepsilon))h(\alpha)}{(\varepsilon + \hat{\alpha}(p - \varepsilon))}$ so that $\hat{H}(x|S, p) = \dfrac{H(x)\varepsilon + (p - \varepsilon)\int_0^x uh(u)\, du}{(\varepsilon + \hat{\alpha}(p - \varepsilon))}$ and thus $\dfrac{\partial \hat{H}(x|S, p)}{\partial p} \leq 0$. This implies that its mean increases in p. The first part of the proposition follows since G^N first order stochastically dominates G. The case of failure is analogous.

While this result may seem to depend on the functional form, Van den Steen (2002) shows that it holds as long as the model captures the allocation of causality to one's own choice of action versus to an exogenous influence. That paper also uses a variation on this model to show that, in a joint project, an agent will tend to attribute a success to his own actions and a failure to the actions of the others.

B. *Overestimation of Control*

Another effect of this mechanism is that an agent may come to overestimate the level of control she has over the outcome. To see this formally, consider the following modification of the model above. Let α's distribution now also depend on the action chosen, i.e., $\alpha_n \sim H_n[0, 1]$. This distribution is again unknown, but agents can form beliefs about it. In particular, let α_n^i be the mean of agent i's prior belief distribution H_n^i about α_n and let it be independently drawn from a distribution without atoms J on $[\underline{\alpha}, \bar{\alpha}]$ with $\underline{\alpha} < \bar{\alpha}$. Let $\mathcal{G} \times \mathcal{J}$ denote the joint distribution of all p_n^i and α_n^i. Assume also that $\varepsilon \in (\underline{p}, \bar{p})$.

PROPOSITION 7: *There exist $\hat{N}$, such that for all $N \geq \hat{N}$, $E_{\mathcal{G}\times\mathcal{J}}[\alpha^1_{Y_1}] > E_{\mathcal{G}\times\mathcal{J}}[\alpha^j_{Y_1}]$.*

PROOF:

I will show that $\lim_{N\to\infty} P[\alpha^1_{Y_1} - \alpha^j_{Y_1} > 0] = 1$. Since α are bounded, this implies the result.

Define the new random variable $\pi_n^i = \varepsilon + \alpha_n^i(p_n^i - \varepsilon) \in [0, 1]$, which is i's expected payoff from choosing a_n, with distribution function denoted K and max supp $K = \bar{\pi} = \bar{p} + \bar{\alpha}(\bar{p} - \varepsilon)$. Agent 1 chooses the action with the highest π_n^1. So $P[\pi^1_{Y_1} \leq x] = P[\pi^1_1 \leq x, \ldots, \pi^1_N \leq x] = K(x)^N \to 0$ for each $x < \bar{\pi}$. It follows that as $N \to \infty$, $\pi^1_{Y_1} \to \bar{\pi}$ with probability 1. This implies that $p^1_{Y_1} \to \bar{p}$ and $\alpha^1_{Y_1} \to \bar{\alpha}$ with probability 1. It follows that $P[\alpha^1_{Y_1} > \alpha^j_{Y_1}] = J(\alpha^1_{Y_1}) \to 1$ with probability 1, which proves the proposition.

In other words, for a sufficient number of distinct alternatives, the agent on average overestimates (according to others) the role that her action choice played.

C. *Overconfidence*

Finally, consider a Bayesian agent who gets a series of data which she combines into a final estimate. Being a Bayesian, she will attach higher weights to data with higher precision. As a consequence, from another agent's perspective, she will attach too much weight to data of which she overestimated the precision. The result is that she will overestimate the precision of the final estimate, i.e., she will be overconfident in the sense of Oskamp (1965). While there are clear similarities, there are also important differences with the current mechanism and some added intricacies that require a separate analysis. I therefore refer to Van den Steen (2004) for the formal analysis.

V. Conclusion

This paper focused on the agent's choice of action, rather than his mental processes, as a source of judgment bias: the bias does not originate in the process of deriving beliefs, but in the choice of which beliefs are relevant, through the choice of action. This process has very different comparative statics than motivational or cognitive theories of judgment biases: it will be strongest when there are many alternatives, when the agent has no prior experience with this particular choice, and when agents tend to disagree more. Moreover, the bias may be influenced by the importance of succeeding. The paper then showed that this mechanism may also play a role in other biases, including asymmetric attributions of causality, overestimation of control, and overconfidence. While the current evidence does not distinguish this theory from others, the theory does have empirical implications that can be verified through experiments and that make such distinction.

The most important contribution of this paper is to suggest a direct link between differing priors and systematic judgment biases. If the theory is confirmed, then it implies that such biases will emerge or increase whenever there is open disagreement.

Appendix: Generalizations

Robustness to More General Distributions

The basic result holds for nearly any distribution of beliefs that is symmetric in the two agents. In particular, let $\mathcal{J}$ be the joint distribution of all p_n^i, for $i = 1, 2$ and $n \in N$. Assume that $\mathcal{J}$ is atomless, symmetric in the two agents, and such that with positive probability the agents (strictly) disagree on the optimal action. Note that this specification allows dependencies not only among the beliefs of the agents, but also among the beliefs over different actions. The following proposition says that, as long as $N > 1$, each agent will consider his own action to be better than that of the other and will be considered to be overoptimistic about his own action by the other.

PROPOSITION A1: *When the number of actions* $N > 1$, $\forall j \neq 1$

- $E_J[\rho_1^1 - \rho_1^j] > 0$
- $\rho_1^1 - \rho_j^1 \geq 0$, *while* $E_J[\rho_1^1 - \rho_j^1] > 0$.

PROOF:

Note first that $\rho_1^1 - \rho_j^1 \geq 0$ by revealed preference. Moreover, by assumption, there is some probability that $\rho_1^1 - \rho_j^1 > 0$. It follows that $E_J[\rho_1^1 - \rho_j^1] > 0$. But now

$$E_J[\rho_1^1 - \rho_1^j] = E_J[\rho_1^1] - E_J[\rho_1^j] = E_J[\rho_1^1] - E_J[\rho_j^1] = E_J[\rho_1^1 - \rho_j^1] > 0$$

where the second equality follows by symmetry. This concludes the proposition.

It is easy to find counterexamples when the joint distribution is not symmetric in the agents. If, for example, $p_n^1 \overset{\text{i.i.d.}}{\sim} G(0, \hat{p})$ and $p_n^2 \overset{\text{i.i.d.}}{\sim} H(\hat{p}, 1)$ then agent 2 will always believe that agent 1 underestimates his chances of success.

A General Utility-Overestimation Proposition

The logic behind the theory in this paper holds in even 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 $\hat{x}_{\mu_i}$ agent i's optimal choice, given these beliefs. Assume that for any belief, such an optimal action exists. This gives the following general proposition.[4]

PROPOSITION A2: $E_{F_{i,j}}[E_i u_i(\hat{x}_{\mu_i}) - E_j u_i(\hat{x}_{\mu_i})] \geq 0$. *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* $\nu \in M$, $\nu \neq \mu$.

$E_i u_i(\hat{x}_{\mu_i}) \geq E_i u_j(\hat{x}_{\mu_j})$. *The inequality will be strict when there is a unique optimal action for agent* i *and* $\hat{x}_{\mu_i} \neq \hat{x}_{\mu_j}$.

PROOF:

First let $\mu_i = \mu$ 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 $\hat{x}_\mu$ and

[4] This proposition was suggested by Bengt Holmstrom.

$\hat{x}_\nu$. $E_i u_i(\hat{x}_{\mu_i}) = \int u(\hat{x}_{\mu_i}, y)\, d\mu_i(y) = \int u(\hat{x}_\mu, y)\, d\mu(y)$ and $E_j u_i(\hat{x}_{\mu_i}) = \int u(\hat{x}_\mu, y)\, d\mu_j(y) = \int u(\hat{x}_\mu, y)\, d\nu(y) \leq \int u(\hat{x}_\nu, y)\, d\nu(y)$ 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, $\int u(\hat{x}_\mu, y)\, d\nu(y) < \int u(\hat{x}_\nu, y)\, d\nu(y)$. The second result follows from revealed preference and the symmetry of the utility functions.

The first part of the proposition says that, on average, an agent relatively overestimates his expected utility in the sense that the agent's estimate of his own expected utility is higher than what others think his expected utility actually is. The second part says that, in this symmetric setting, each agent expects to have higher utility than any other agent.

REFERENCES

Baumeister, Roy F. "The Self," in Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey, eds., *Handbook of Social Psychology*. New York: McGraw-Hill, 1998, pp. 680–740.

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

Blackwell, David and Dubins, Lester E. "Merging of Opinions with Increasing Information." *Annals of Mathematical Statistics*, September 1962, *38*(3), pp. 882–86.

Bodner, Ronit and Prelec, Drazen. "Self-Signaling and Diagnostic Utility in Everyday Decision Making," in Isabelle Brocas and Juan D. Carrillo, eds., *The psychology of economic decisions, volume 1*: *Rationality and well-being*. Oxford: Oxford University Press, 2003, pp. 105–23.

Brocas, Isabelle and Carrillo, Juan D. "Are We All Better Drivers Than Average? Self-Perception and Biased Behavior." Working paper, Columbia Business School, 2002.

Brown, Keith C. "A Note on the Apparent Bias of Net Revenue Estimates for Capital Investment Projects." *Journal of Finance*, September 1974, *29*(4), pp. 1215–216.

Brunnermeier, Markus K. and Parker, Jonathan A. "Optimal Expectations." Working paper, Princeton University, 2003.

Capen, Edward; Clapp, Robert and Campbell, William. "Competitive Bidding in High-Risk Situations." *Journal of Petroleum Technology*, June 1971, *23*(6), pp. 641–53.

Compte, Olivier. "The Winner's Curse with Independent Private Values." Working paper, CERAS-ENPC, 2002.

Compte, Olivier and Postlewaite, Andrew. "Confidence Enhanced Performance." PIER Working Paper No. 03-009, 2003.

Duval, Thomas S. and Silvia, Paul J. "Self-Awareness, Probability of Improvement, and the Self-serving Bias." *Journal of Personality and Social Psychology*, January 2002, *82*(1), pp. 49–61.

Fiske, Susan and Taylor, Shelley. *Social cognition*. New York: McGraw-Hill, 1991.

Harrison, J. Richard and March, James G. "Decision Making and Postdecision Surprises." *Administrative Science Quarterly*, March 1984, *29*(1), pp. 26–42.

Harrison, Michael and Kreps, David M. "Speculative Investor Behavior in a Stock Market with Heterogenous Expectations." *Quarterly Journal of Economics*, May 1978, *92*(2), pp. 323–36.

Harsanyi, John C. "Games with Incomplete Information Played by 'Bayesian' Players, I–III, Part III. The Basic Probability Distribution of the Game." *Management Science*, March 1968, *14*(7), pp. 486–502.

Hvide, Hans K. "Pragmatic Beliefs and Overconfidence." *Journal of Economic Behavior and Organization*, May 2002, *48*(1), pp. 15–28.

Ickes, William J. "A Basic Paradigm for the Study of Personality, Roles, and Social Behavior," in William J. Ickes and Eric S. Knowles, eds., *Personality, roles, and social behavior*. New York: Springer-Verlag, 1982, pp. 305–41.

Kahneman, Daniel and Tversky, Amos. "On the Psychology of Prediction." *Psychological Review*, July 1973, *80*(4), pp. 237–51.

Kunda, Ziv. "The Case for Motivated Reasoning." *Psychological Bulletin*, November 1990, *108*(3), pp. 480–98.

Langer, Ellen J. "The Illusion of Control." *Journal of Personality and Social Psychology*, August 1975, *32*(2), pp. 311–28.

Lazear, Edward. "The Peter Principle: A Theory of Decline." *Journal of Political Economy*, February 2004, *112*(1), pp. S141–63.

Miller, Dale T. and Ross, Michael. "Self-serving Biases in the Attribution of Causality: Fact or Fiction?" *Psychological Bulletin*, March 1975, *82*(2), pp. 213–25.

Mischel, Walter. "The Interaction of Person and Situation," in David Magnusson and Norman S. Endler, eds., *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977, pp. 333–52.

Morris, Stephen. "The Common Prior Assumption in Economic Theory." *Economics and Philosophy*, October 1995, *11*(2), pp. 227–53.

______. "Speculative Investor Behavior and Learning." *Quarterly Journal of Economics*, November 1996, *111*(4), pp. 1111–33.

Oskamp, Stuart. "Overconfidence in Case-Study Judgments." *Journal of Consulting Psychology*, June 1965, *29*(3), pp. 261–65.

Peter, Laurence and Hull, Raymond. *The Peter Principle*. New York: W. Morrow, 1969.

Rabin, Matthew and Schrag, Joel. "First Impressions Matter: A Model of Confirmation Bias." *Quarterly Journal of Economics*, February 1999, *114*(1), pp. 37–82.

Ross, Michael and Sicoly, Fiore. "Egocentric Biases in Availability and Attribution." *Journal of Personality and Social Psychology*, March 1979, *37*(3), pp. 322–36.

Rothschild, Michael. "A Two-Armed Bandit Theory of Market Pricing." *Journal of Economic Theory*, October 1974, *9*(2), pp. 185–202.

Santos-Pinto, Luis and Sobel, Joel. "A Model of Positive Self-Image in Subjective Assessments." Working paper, UCSD, 2003.

Snyder, Mark and Ickes, William J. "Personality and Social Behavior," in Gardner Lindzey and Elliott Aronson, eds., *Handbook of social psychology*, Vol. 2, 3rd Ed. New York: Random House, 1985, pp. 883–947.

Van den Steen, Eric J. "Essays on the Managerial Implications of Differing Priors." Ph.D. dissertation, Stanford University, 2001.

______. "Skill or Luck: Biases of Rational Agent." MIT-Sloan Working Paper No. 4255-02, 2002.

______. "Overconfidence by Rational Agents." Working paper, MIT, 2004.

Weinstein, Neil D. "Unrealistic Optimism about Future Life Events." *Journal of Personality and Social Psychology*, November 1980, *39*(5), pp. 806–20.

______. "Unrealistic Optimism about Susceptibility to Health Problems." *Journal of Behavioral Medicine*, 1982, *5*(4), pp. 441–60.

______. "Why It Won't Happen to Me: Perceptions of Risk Factors and Susceptibility." *Health Psychology*, 1984, *3*(5), pp. 431–57.

Zábojník, Ján. "A Model of Rational Bias in Self-Assessments." *Economic Theory*, February 2004, *23*(2), pp. 259–82.

Zuckerman, Miron. "Attribution of Success and Failure Revisited, or: The Motivational Bias Is Alive and Well in Attribution Theory." *Journal of Personality*, June 1979, *47*(2), pp. 245–87.