Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 04-023

"Confidence-Enhanced Performance"

by

Olivier Compte and Andrew Postlewaite

http://ssrn.com/abstract=549162

Confidence-Enhanced Performance*

Olivier Compte
CERAS

Andrew Postlewaite
University of Pennsylvania

May, 2001
This version: May, 2003

Abstract

There is ample evidence that emotions affect performance. Positive emotions can improve performance, while negative ones may diminish it. For example, the fears induced by the possibility of failure or of negative evaluations have physiological consequences (shaking, loss of concentration) that may impair performance in sports, on stage or at school.

There is also ample evidence that individuals have distorted recollection of past events, and distorted attributions of the causes of successes of failures. Recollection of good events or successes is typically easier than recollection of bad ones or failures. Successes tend to be attributed to intrinsic aptitudes or own effort, while failures are attributed to bad luck. In addition, these attributions are often reversed when judging the performance of others.

The objective of this paper is to incorporate the first phenomenon above into an otherwise standard decision theoretic model, and show that in a world where performance depends on emotions, biases in information processing enhance welfare.

1. Introduction

A person's performance can be affected by his or her psychological state. One common form in which this is manifested is "choking", which is a physical response to a perceived psychological situation, usually fear of not performing well and failing. The fear response physically affects the individual: breathing may get short and shallow and muscles tighten. This may be the response to fear of performing badly on tests, on the golf course or when speaking in public; in other words, it may occur any time that an individual's performance at some activity matters. The fear of performing badly is

*Much of this work was done while Postlewaite was a visitor at CERAS. Their support is gratefully acknowledged, as is the support of the National Science Foundation. We thank Doug Bernheim and two anonymous referees for suggestions that greatly improved the paper.

widespread: it is said that people put public speaking above death in rankings of fears.[1] The response is ironic in that exactly when performance matters, the physiological response may compromise performance. Whether it's a lawyer presenting a case to the Supreme Court, a professional golfer approaching a game-winning putt or an academic giving a job market seminar, the probability of success diminishes if legs are trembling, hands are shaking or breathing is difficult.[2]

The fear that triggers the physiological response in a person faced with a particular activity is often related to past events similar to that activity, and is likely to be especially strong if the person has performed less than optimally at those times. Knowing that one has failed frequently in the past at some task may make succeeding at that task in the future difficult. Stated succinctly, the likelihood of an individual's succeeding at a task may not be independent of his or her *beliefs* about the likelihood of success.

Neoclassical decision theory does not accommodate this possibility; individual's choose whether or not to undertake an activity depending on the expected utility (or some other criterion) of doing so, where it is assumed that the probability distribution over the outcomes is exogenously given. We depart from neoclassical decision theory by incorporating a "performance technology" whereby a person's history of successes and failures at an activity affect the probability of success in future attempts. We show that the presence of this type of interaction may induce biases in information processing similar to biases that have been identified by psychologists. These biases are sometimes seen by economists as anomalies that are difficult to reconcile with rational decision making. Rather than being a liability, in our model these biases increase an individual's welfare.

We review related literature in the next section, work in psychology describing the effect of psychological state on performance and systematic biases in people's decision making. We then present the basic model, including the possibility that an individual's perceptions can be biased, in section 3, and show in the following section the optimality of biased perceptions. We conclude with a discussion section that includes related work.

2. Related work in psychology

Our main point is that an individual's psychological state affects his or her performance, and that as a consequence, it is optimal for people to have biased perceptions. We will review briefly some work done by psychologists that supports the view that an

[1] From the Speech anxiety website at Rochester University;
http://www.acd.roch.edu/spchcom/anxiety.htm.

[2] There may be, of course, physiological responses that may be beneficial, such as a surge of adrenaline that enhances success in some athletic pursuits. Our focus in this paper is on those responses that diminish performance.

individual's psychological state affects his or her performance, as well as the evidence concerning biased perceptions.

2.1. Psychological state affecting performance

Psychological state can affect performance in different ways. Steele and Aronson (1995) provide evidence that stress can impair performance. Blacks and whites were given a series of difficult verbal GRE problems to solve under two conditions: the first was described as diagnostic, to be interpreted as evaluating the individual taking the test, while the purpose of the second test was described as determining how individuals solved problems. Steele and Aronson interpret the diagnostic condition as putting Black subjects at risk of fulfilling the racial stereotype about their intellectual ability, causing self-doubt, anxiety, etc., about living up to this stereotype. Blacks performed more poorly when stress is induced (the diagnostic condition) than under the neutral condition.[3]

Aronson, Lustina, Good, and Keough (1999) demonstrate a similar effect in white males. Math-proficient white males did measurably worse on a math test when they were explicitly told prior to the test that there was a stereotype that Asian students outperform Caucasian students in mathematical areas than did similar students to whom this statement was not made.

Taylor and Brown (1988) survey a considerable research that suggests that overly positive self-evaluation, exaggerated perceptions of mastery and unrealistic optimism are characteristic of normal individuals, and that moreover, these illusions appear to promote productive and creative work.

There are other psychological states that can affect performance. Ellis et al. (1997) showed that mood affected subjects' ability to detect contradictory statements in written passages. They induced either a neutral or a depressed mood by having participants read aloud a sequence of twenty-five self referent statements. An example of a depressed statement was "I feel a little down today," and an example of a neutral statement was "Sante Fe is the capital of New Mexico."[4] They found that those individuals in whom a depressed mood was induced were consistently impaired in detecting contradictions in prose passages. McKenna and Lewis (1994) similarly demonstrated that induced mood affected articulation. Depressed or elated moods were induced in participants, who were then asked to count aloud to 50 as quickly as possible. Performance was retarded in the depressed group.

Physical reaction time was shown to be affected by induced mood by Brand, Verspui and Oving (1997). Subjects were randomly assigned to two groups with positive mood induced in one and negative mood in the other. Positive mood was induced by showing the (Dutch) subjects a seven minute video consisting of fragments of the 1988 European

[3] See also Steele and Aronson (1998) for a closely related experiment.

[4] See Teasdale and Russell (1983) for a fuller description and discussion of this type of mood induction.

soccer championship in which the Dutch team won the title, followed by a two minute fragment of the movie "Beethoven" including a few scenes of a puppy dog. Negative mood was induced by showing subjects an eight minute fragment of the film "Faces of Death" consisting of a live recorded execution (by electric chair) of a criminal.[5] Subjects with positive induced mood showed faster response times than did subjects with negative induced mood.

Baker et al. (1997) also induce elated and depressed moods in subjects,[6] and show that induced mood affects subjects' performance on a verbal fluency task. In addition, Baker et al. measure regional cerebral blood flow using Positron Emission Tomography (PET). They find that induced mood is associated with activation of areas of the brain associated with the experience of emotion. This last finding is of particular interest in that it points to demonstrable physiological effects of mood.

In our formal model below, we assume that the psychological state that affects performance on a particular task is associated with related tasks done in the past. We should point out that only the first half (approximately) of the work surveyed above can be considered as supporting this. We include the remainder because they provide evidence of the broader point that we think important – that the probability that an individual will be successful at a task should *not* be taken as always being independent of psychological factors.

2.2. Biased perception

Psychologists have compiled ample evidence that people have biased perceptions, and in particular, that in comparison to others, individuals systematically evaluate themselves more highly than others do. Guthrie, Rachlinski and Wistrich (2001) distributed a questionnaire to 168 federal magistrate judges as part of the Federal Judicial Center's Workshop for Magistrate Judges II in New Orleans in November 1999. The respondents were assured that they could not be identified from their questionnaires, and were told that they could indicate on their questionnaire if they preferred their answers not to be included in the research project. One of the 168 judges chose not to have his or her answers included.

To test the relationship of the judges' estimates of their ability relative to other judges, the judges were asked the following question to estimate their reversal rates on appeal: "United States magistrate judges are rarely overturned on appeal, but it does

[5] Subjects in this group were explicitly told in advance that they were free to stop the video if the emotional impact of the fragments would become too strong. Seven of the twenty subjects made use of this possibility about halfway through the video.

[6] Depressed, neutral and elated states are induced by having subjects listen to extracts of Prokofiev's "Russia under the Mongolian Yoke" at half speed in the depressed state, "Stressbusters," a recording of popular classics in the neutral condition, and Delibes' "Coppelia" in the elated condition.

occur. If we were to rank all of the magistrate judges currently in this room according to the rate at which their decisions have been overturned during their careers, [what] would your rate be?" The judges were then asked to place themselves into the quartile corresponding to their respective reversal rates: highest ($> 75\%$), second highest ($> 50\%$), third highest ($> 25\%$), or lowest.

The judges answers are very interesting: 56.1% put themselves into the lowest quartile, 31.6% into the second lowest quartile, 7.7% in the second highest quartile and 4.5% in the highest quartile. In other words, nearly 90% thought themselves above average.

These judges are not alone in overestimating their abilities relative to others. People routinely overestimate themselves relative to others in driving, (Svenson (1981)), the likelihood their marriage will succeed (Baker and Emery (1993)), and the likelihood that they will have higher salaries and fewer health problems than others in the future (Weinstein (1980)). Ross and Sicoly (1979) report that when married couples are asked to estimate the percentage of household tasks they perform, their estimates typically add up to far more than 100%.

People do not only overestimate their likelihood of success relative to other people; they overestimate the likelihood of success in situations not involving others as well. When subjects are asked questions, and asked the likelihood that they are correct along with their answers their "hit rate" is typically 60% when they are 90% certain (see, e.g., Fischoff, Slovic and Lichtenstein (1977) and Lichtenstein, Fischoff and Phillips (1982)).

Biases in judging the relative likelihood of particular events are frequent, and many scholars have attempted to trace the source of these biases. One often mentioned source is the over-utilization of simple heuristics, such as the *availability heuristic* (Tversky and Kahneman (1973)): in assessing the likelihood of particular events, people are often influenced by the availability of similar events in their past experience. For example, a worker who is often in contact with jobless individuals for example because he is jobless himself, would typically overestimate the rate of unemployment; similarly, an employed worker would typically underestimate the rate of unemployment. (See Nisbett and Ross (1980), page 19). In the same vein, biases in self-evaluations may then be the result of some particular past experiences being more readily available than others: if good experiences are more easily remembered than bad ones, or if failures tend to be disregarded or attributed to atypical circumstances, people will tend to have overly optimistic self-evaluations (see Seligman (1990) for more details on the connection between attributional styles and optimism).

Economists often see these biases as "shortcomings" of judgement or pathologies that can only lower the welfare of an individual, and should be corrected. As mentioned in the introduction, such biases will emerge in our model naturally as welfare enhancing.

3. Basic model

We consider an agent who faces a sequence of decisions of whether or not to undertake a risky activity. This activity may either be a success or a failure. We have in mind a situation as follows. Consider a lawyer who is faced with a sequence of cases that he may accept or decline. Accepting a case is a risky prospect as he may win or lose the case. The lawyer will receive a payoff of 1 in case of success, and a payoff normalized to 0 in case of failure.

Our primary departure from conventional decision theory is that we assume that the lawyer's probability of winning the case depends on his confidence:[7] if he is unsure about his abilities, or anxious, his arguments will have less chance of convincing the jury. To capture the idea that confidence affects performance in the simplest way, we identify these feeling of anxiety or self assurance with the lawyer's perception of success in previous cases. A lawyer who is more confident about his chances of success will be assumed to have greater chances of succeeding than a lawyer who is not confident.

In what follows, we present two main building blocks of our model: First we describe the risky activity, the true prospects of the agent if he undertakes it, and the effect of confidence. Second, we describe the decision process followed by the agent.

3.1. The risky activity

We start by modelling confidence and its effect on performance. We shall make two assumptions. First, higher confidence will translate into higher probability of success. Second, confidence will depend on the agent's perception on how successful he has been in the past.

Formally, we denote ρ the probability of success, and we take the parameter $\kappa \in (0, 1]$ as a measure of the agent's confidence. We assume that

$$\rho = \kappa\rho_0,$$

where $\rho_0 \in (0, 1)$. The probability ρ_0 depends solely on the characteristics of the project being undertaken. Absent any psychological considerations, the agent would succeed with probability ρ_0. Lack of confidence reduces performance.[8]

Note that we do not model the process by which performance is reduced. It is standard to assume that performance should only depend on the agent' actions while undertaking the project. Hence one might object that we need to explain how confidence alters the decisions made by the agent while undertaking the project. We do

[7] We use the term "confidence" in a broad sense, and mean it to include feelings of assuredness and lack of anxiety.

[8] Note that we do not assume that the agent knows ρ (nor κ or ρ_0). We discuss this issue in Subsection 3.2, when we describe how the agent forms beliefs about his chance of success.

not model these decisions however. We have in mind that the agent does not have full conscious control over all the actions that need to be undertaken, and that the agent's psychological state of mind has an effect on these actions that is beyond the agent's conscious control.

We now turn to modelling how confidence (or lack of confidence) arises. Various factors may contribute to decreasing confidence, such as the remembrance of past failures, or the agent's perception of how likely he is to succeed. In the basic version of our model, we shall assume confidence depends solely on the agent's perception of the empirical frequency of past success.

Formally, we denote by s (respectively f) the number of successes (respectively failures) that the agent recalls. We explicitly allow the number of successes and failures that the agent recalls to deviate from the actual number.[9] We will explain below how this data is generated. We define the *perceived empirical frequency of success* as

$$\varphi = \frac{s}{s+f},$$

and we assume that confidence is a smooth and increasing function of φ, that is,

$$\kappa = \kappa(\varphi),$$

where $\kappa' > 0$, $\kappa(0) > 0$ and, without loss of generality, $\kappa(1) = 1$.[10]

Combining the two assumptions above, we may view the probability of success as a function of the agent's perception φ, and define the performance function $\boldsymbol{\rho}$ as follows:

$$\boldsymbol{\rho}(\varphi) = \kappa(\varphi)\rho_0$$

Lastly, we shall assume that the effect of confidence on performance is not too strong, that is, that

$$\boldsymbol{\rho}'(\varphi) < 1.$$

This assumption is made mostly for technical convenience. It will be discussed below and in the Appendix. We draw the function $\boldsymbol{\rho}$ in Figure 1 below. The positive slope captures the idea that high (respectively low) perceptions affect performance positively (respectively negatively).

---

[9] Our modelling here is similar to that of Rabin and Schrag (1999); we discuss the relationship between the models in the last section.

[10] When no data is available, we set $\kappa = 1$. Our results will not depend on that assumption however.

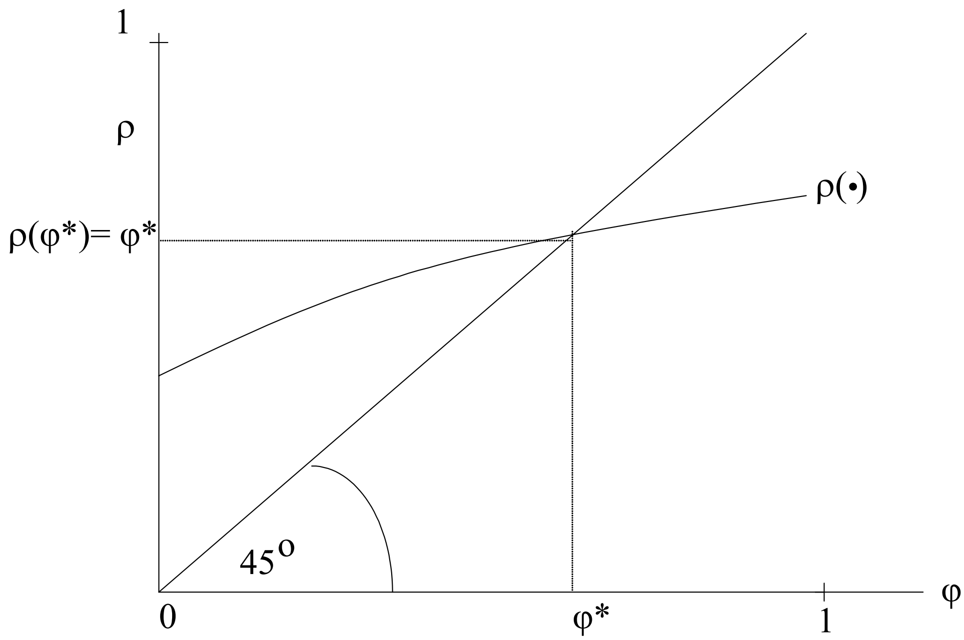

Figure 1

Among the possible perceptions φ that an agent may have about the frequency of success, there is one that will play a focal role. Consider the (unique) perception φ^* such that[11]

$$\varphi^* = \boldsymbol{\rho}(\varphi^*).$$

Whenever the agent has a perception φ that is below φ^*, the objective probability of success $\rho(\varphi)$ exceeds his perception φ. Similarly whenever the agent holds a perception φ that is above φ^*, the objective probability of success is below his perception. At φ^*, the agent's perception is equal to the objective probability of success. An agent who holds this perception would not find his average experience at variance with his perception.

We now turn to how the numbers s and f are generated. Obviously, there should be some close relationship between true past outcomes and what the agent perceives or recalls. In many contexts however, the psychology literature has identified various factors such as generation of attributions or inferences, memorability, perceptual salience, vividness, that may affect what the agent perceives, records and recalls, hence the data that is available to the agent.

[11] There is a unique ϕ^* satisfying this equation due to our assumption that $\boldsymbol{\rho}' \in [0, 1)$.

To illustrate, for example, the role of attributions, consider the case of our lawyer. When he loses, he may think that this was due to the fact that the defendant was a member of a minority group while the jury was all white: the activity was a failure, but the reasons for the failure are seen as atypical and not likely to arise in the future. He may then disregard the event when evaluating his overall success. In extreme cases, he may even convince himself that had the circumstances been "normal", he would have won the case and record the event as a success in his mind.

This basic notion is not new; there is evidence in the psychology literature that people tend to attribute positive experiences to things that are permanent and to attribute negative experiences to transient effects.[12] If one gets a paper accepted by a journal, one attributes this to being a good economist, while rejections are attributed to the referee not understanding the obviously important point being made. With such attribution, successes are likely to be perceived as predicting further successes, while failures have no predictive content.

We now formalize these ideas. We assume that after undertaking the activity, if the outcome is a failure, the agent, with probability $\gamma \in [0, 1)$, attributes the outcome to atypical circumstances. Events that have been labelled atypical are not recorded. When $\gamma = 0$, perceptions are correct, and the true and perceived frequencies of success coincide. When $\gamma > 0$, perceptions are biased: the agent overestimates the frequency of past success.

To assess precisely how the perception bias γ affects the agent's estimate of his past success rate, define the function $\psi^\gamma : [0, 1] \to [0, 1]$ as

$$\psi^\gamma(\rho) = \frac{\rho}{\rho + (1-\gamma)(1-\rho)}.$$

$\psi^\gamma(\rho)$ is essentially the proportion of recorded events that are recorded as success when the true frequency of success is ρ and γ is the perception bias. True and perceived frequencies of success are random variables. Yet in the long run, as will be shown in the appendix, their distributions will be concentrated around a single value, and a true frequency of success ρ will translate into a perceived frequency of success

$$\varphi = \psi^\gamma(\rho)$$

Equivalently, for the agent to perceive a success rate of φ in the long run, the true long-run frequency of success has to be equal to $(\psi^\gamma)^{-1}(\varphi)$. Figure 2 illustrates such a function with $\gamma > 0$. Note that with our assumptions, perceptions are optimistic: they always exceed the true frequency of past success, that is, $\psi^\gamma(\rho) > \rho$.

[12] See, e.g., Seligman (1990).

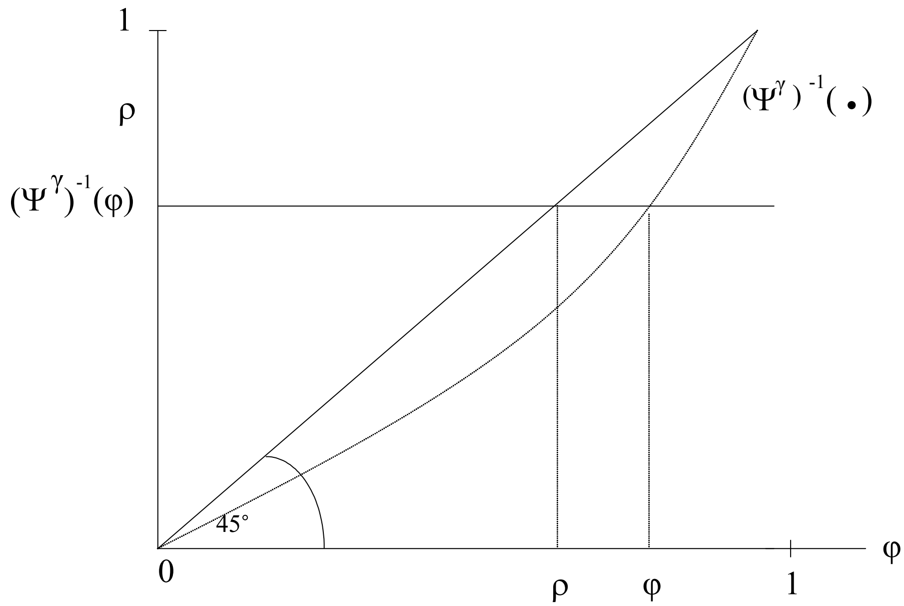

Figure 2

3.2. Decision process

As described in the outline of our model above, the agent faces a sequence of decisions of whether or not to undertake a risky activity. We assume that undertaking the activity entails a cost c, possibly because of a forgone opportunity. For example, in the lawyer example above, there might always be available some riskless alternative to taking on a new case, such as drawing up a will. We assume the cost c is stochastic and takes values in $[0, 1]$. We further assume that the random variables $\{c_t\}_{t=1}^{\infty}$, the costs at each time t, are independent, that the support of the random variables is $[0, 1]$, and that at the time the agent chooses whether to undertake the activity at t, he knows the realization of c_t.

To evaluate whether undertaking the project is worthwhile, the agent forms a belief p about whether he will succeed, and we assume this belief is based on the data he recollects. This can be viewed as a formalization of the *availability heuristic*: only the outcomes that are recollected are *available* to the agent. We do not model however the fine details of how recollections of past successes and failures are mapped into the agent's belief about the probability of success on the next try; rather, we model this process as a function

$$p = \beta(s, f).$$

One can think of this function as capturing the dynamic process by which an agent who is initially unsure about ρ_0 would update his initial beliefs as he accumulates experience.[13] Alternatively, one might imagine an agent who follows a simple rule of thumb in forming beliefs. The only restrictions we place on the function β are the following.[14]

(i) $\forall s, f \geq 0$, $0 < \beta(s, f) < 1$

(ii) There exists $A > 0$ such that $\forall s, f > 0$, $\mid \beta(s, f) - \frac{s}{s+f} \mid \leq A/(s + f)$

The first assumption is a statement that beliefs must lie between 0 and 1. The second assumption is an "asymptotic consistency" condition that rules out belief formation processes for which there is a permanent divergence between the agent's perceived successes and failures and his beliefs: when the number of recalled past outcomes is large, the belief of the agent must approach the perceived empirical frequency of past successes.

Under these conditions, when the data is not biased, beliefs are correct is the long run. We insist on the same conditions holding even when the data is biased, because our view is that the agent is ignorant of the fact that the process by which this data is generated might be biased, and lead him to biased beliefs.[15]

Having formed a belief p_t about his chance of success at date t if he undertakes the activity, the agent compares the expected payoff from undertaking the activity to the cost of undertaking it. That is, the agent undertakes the activity if and only if[16]

$$p_t \geq c_t.$$

Finally, we define the expected payoff $v(p, \varphi)$ to the agent who has a belief p, a

[13] Depending on the priors of the agent and on what the agent is trying to estimate, Bayesian learning explanations for the formation of beliefs might depend on the sequence of successes and failures. To facilitate exposition, we have chosen to describe the function β as a function of the aggregate numbers of recalled successes and failures only. One could alternatively consider more complicated functions determining the agent's belief about the probability of success that depend on the particular sequencing of recalled successes and failures.

[14] These restrictions are in addition to the assumption that it is only the aggregate numbers of recalled successes and failures that matter.

[15] We will return to this issue in the Discussion Section, and examine the case where the agent is aware of the fact that his data may be biased, and attempt to correct his beliefs accordingly.

[16] As mentioned above, we have not assumed that the agent knows ρ_0, hence he might wish to experiment. We could allow for more sophisticated rules of behavior, accounting for the fact that the agent might want to experiment for a while, and undertake the activity when $p_t \geq \mu_t c_t$ with $\mu_t \leq 1$. Our results would easily carry over to this case, so long as with probability 1, there is no experimentation in the long run: $\mu_t \to 1$.

perception φ, and who does not yet know the realization of the cost of the activity:[17]

$$v(p, \varphi) = \Pr\{p \geq c\} E[\boldsymbol{\rho}(\varphi) - c \mid p \geq c].$$

This completes the description of the model. The key parameters of the model are the technology $\boldsymbol{\rho}(\cdot)$, the belief function β, and the perception bias γ. Each triplet (ρ, β, γ) induces a probability distribution over beliefs, perceptions, decisions and outcomes.[18] We are interested in the limit distribution and the agent's expected gain under that limit distribution when the perception bias is γ. We will verify in the Appendix that this limit distribution is well-defined. Formally, let

$$V_t(\gamma) = E_{(\rho,\beta,\gamma)} v(p_t, \varphi_t).$$

We are interested in the long-run payoff defined by

$$V_\infty(\gamma) = \lim_{t \to \infty} V_t(\gamma).$$

We will show that $V'_\infty(0) > 0$, that is, that according to this criterion some bias in perceptions will make the decision maker better off.

Note that our focus on long-run payoffs implies that we do not analyze what effect biased beliefs have on the decision maker's welfare in the transition period leading up to steady state. However, that biases in perceptions might be welfare enhancing in transition would not be surprising: in our model, the decision maker typically starts with erroneous beliefs,[19] and hence make poor decisions in transition.[20] Biases in perceptions could induce better decisions, because they may induce a quicker move towards beliefs that would be closer to true probability of success. We discuss this further in section 4.3.

4. The optimality of biased perceptions

4.1. A benchmark: the cost of biased perceptions

We begin with the standard case where confidence does not affect performance, that is, the case in which $\kappa = 1$, independently of φ. Our objective here is to highlight the

[17] Note that this expected utility is from the perspective of an outside observer, since it is calculated with the true probability of success, $\rho(p)$. From the agent's point of view, the expected value is

$$\Pr\{p \geq c\} E[p - c \mid p \geq c].$$

.

[18] The distribution at date t also depends on the initial condition, that is, the value of confidence when no data is available (which has been set equal to 1; see footnote 10). Under the assumption that $\rho(\cdot)$ is not too steep, however, the limit distribution is independent of initial condition.

[19] Our model puts few contraints on beliefs, except in the long run. In the long run if his perceptions are unbiased, our decision maker will asymptotically have correct beliefs.

[20] For example, an initially pessimistic agent undertakes the project too little.

potential cost associated with biased perceptions.

The probability of success is equal to ρ_0. By the law of large numbers, as the number of instances where the activity is undertaken increases, the frequency of success gets close to ρ_0 (with probability close to 1). When perceptions are correct ($\gamma = 0$), the true and perceived frequencies of past success coincide. Hence, given our assumption (iii) concerning β, the agent's belief must converge to ρ_0 as well. It follows that

$$V_\infty(0) = \int_{\rho_0 \geq c} (\rho_0 - c) g(c) dc$$

where $g(\cdot)$ is the density function for the random cost.

How do biased perceptions affect payoffs? With biased perceptions ($\gamma > 0$), the true frequency of success still gets close to ρ_0 (with probability close to 1). The perceived frequency of success however will get close to $\psi^\gamma(\rho_0)$, and so also will his belief about his chance of success. The agent will thus decide to undertake the project whenever $c \leq \psi^\gamma(\rho_0)$. It follows that

$$V_\infty(\gamma) = \int_{\psi^\gamma(\rho_0) \geq c} (\rho_0 - c) g(c) dc.$$

The only effect of the perception bias γ is to change the circumstances under which the activity is undertaken. When $\gamma > 0$, there are events for which

$$\psi^\gamma(\rho_0) > c > \rho_0.$$

In these events, the agent undertakes the activity when he should not do so (since they have negative expected payoff with respect to the true probability of success), and in the other events, he is taking the correct decision. So the agent's (true) welfare would be higher if he had correct perceptions. This is essentially the argument in classical decision theory that biasing perceptions can only harm agents.

4.2. Confidence enhanced performance

When confidence affects performance, it is no longer true that correct perceptions maximize long-term payoffs. It will still be the case that agents with biased perceptions will have overly optimistic beliefs, and consequently be induced to undertake the activity in events where they should not have. However, on those projects they undertake, their optimism leads to higher performance, that is, they have higher probability of success. We will compare these two effects and show that having some degree of optimism is preferable to correct perceptions.

The key observation (see the appendix) is that in the long run, the perceived frequency of success tends to be concentrated around a single value, say φ_∞, and the possible values of φ_∞ are therefore easy to characterize. The realized frequency of success $\rho_t = \boldsymbol{\rho}(\varphi_t)$ is with high probability near $\rho_\infty = \boldsymbol{\rho}(\varphi_\infty)$. True and perceived frequencies of success are thus concentrated around ρ_∞ and $\psi^\gamma(\rho_\infty)$ respectively. The only possible candidates for φ_∞ must therefore satisfy:

$$\psi^\gamma(\boldsymbol{\rho}(\varphi_\infty)) = \varphi_\infty. \tag{4.1}$$

Below, we will restrict our attention to the case where there is a unique such value. Under the assumption $\boldsymbol{\rho}' < 1$, this is necessarily the case when γ is not too large.[21] We will discuss in the Appendix the more general case where Equation (4.1) may have several solutions.

Figure 3 illustrates geometrically φ_∞ for an agent with biased perceptions (and consequently, optimistic beliefs).

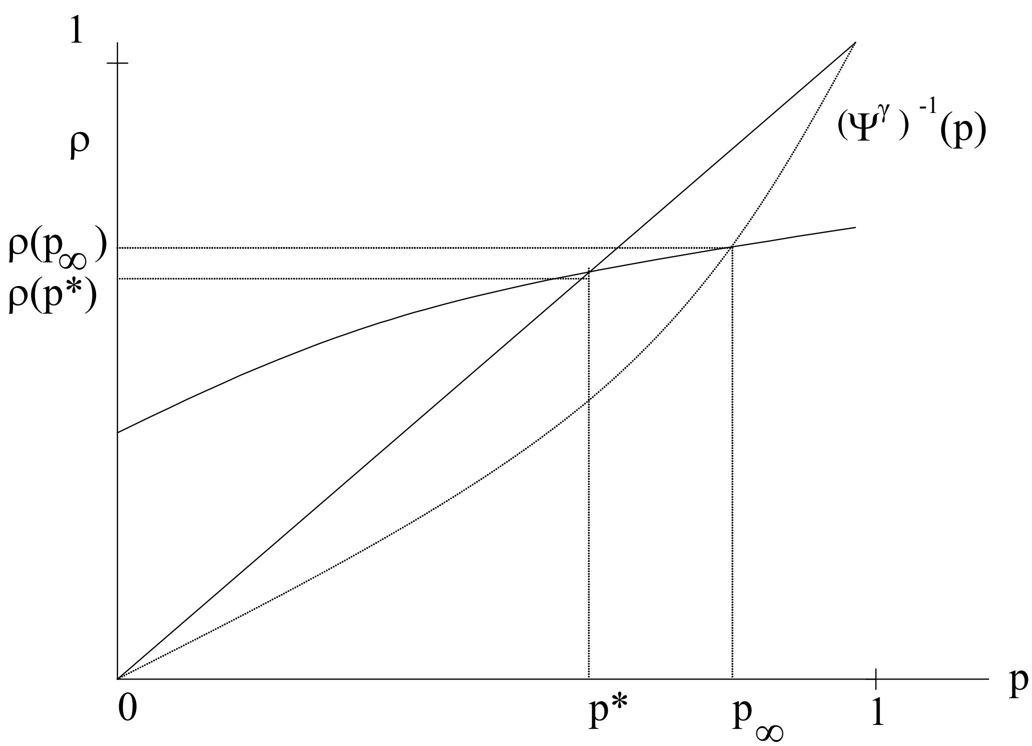

Figure 3

[21] This is because for γ not too large, $(\psi^\gamma \circ \boldsymbol{\rho})' < 1$.

Note that in the case that perceptions are correct, φ_∞ coincides with the rational belief φ^* defined earlier, but that for an agent for whom $\rho'(\cdot) > 0$,

$$\varphi_\infty > \boldsymbol{\rho}(\varphi_\infty) > \boldsymbol{\rho}(\varphi^*) = \varphi^*.$$

In the long-run, the agent with positive perception bias thus has higher performance than an agent with correct perceptions ($\rho(\varphi_\infty) > \rho(\varphi^*)$), but he is overly optimistic about his chances of success ($\varphi_\infty > \rho(\varphi_\infty)$).

Turning to the long-run payoff, we obtain:

$$V_\infty(\gamma) = \int_{\varphi_\infty \geq c} (\boldsymbol{\rho}(\varphi_\infty) - c) g(c) dc.$$

As a benchmark, the long-run payoff when perceptions are correct is equal to

$$V_\infty(0) = \int_{\varphi^* \geq c} (\varphi^* - c) g(c) dc.$$

To understand how $V_\infty(\gamma)$ compares to $V_\infty(0)$, we write $V_\infty(\gamma) - V_\infty(0)$ as the sum of three terms:

$$\begin{aligned} V_\infty(\gamma) - V_\infty(0) &= \int_{c \leq \varphi^*} (\boldsymbol{\rho}(\varphi_\infty) - \boldsymbol{\rho}(\varphi^*)) g(c) dc + \int_{\varphi^*}^{\boldsymbol{\rho}(\varphi_\infty)} (\boldsymbol{\rho}(\varphi_\infty) - c) g(c) dc \\ &\quad + \int_{\boldsymbol{\rho}(\varphi_\infty)}^{\varphi_\infty} (\boldsymbol{\rho}(\varphi_\infty) - c) g(c) dc \end{aligned} \tag{4.2}$$

The first term is positive and corresponds to the increase in performance that arises due to optimism for the activities that would have been undertaken even if perceptions had been correct. The second term is positive: it corresponds to the realizations of costs for which the activity is profitable to the agent only because he is optimistic. Finally, the third term is negative: it corresponds to the realizations of costs for which the agent should not have undertaken the activity and undertakes it because he is optimistic. The shaded regions in figure 4 represent these three terms when c is uniformly distributed.

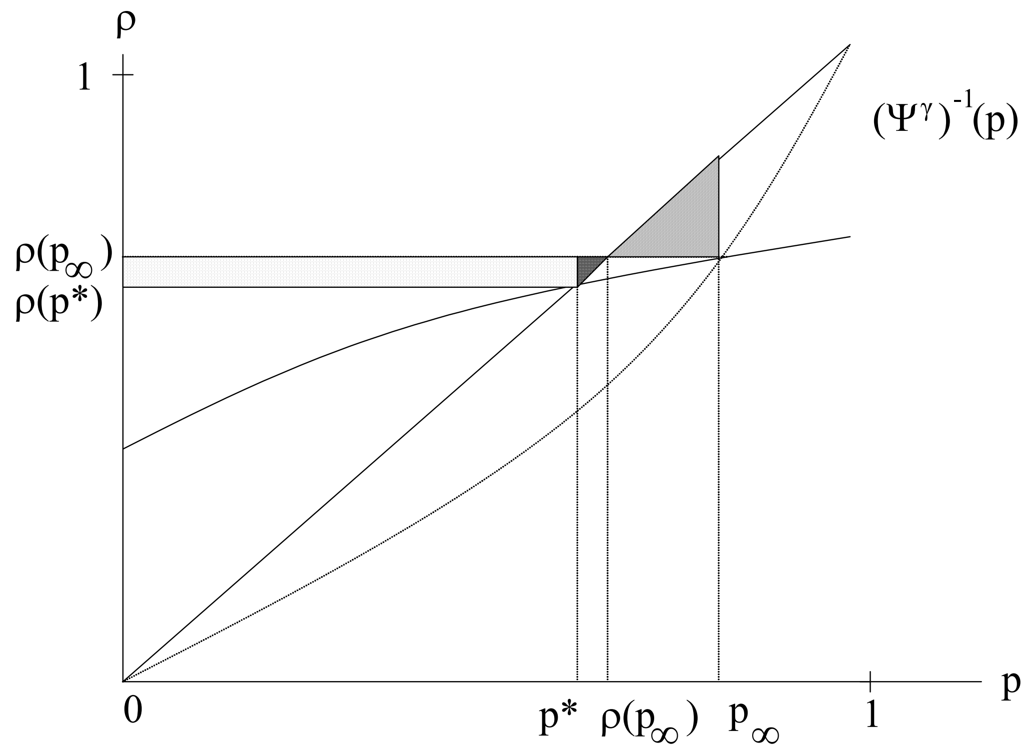

Figure 4

One implication of equation (4.2) is that when confidence positively affects performance ($\rho' > 0$), some degree of optimism always generates higher long-run payoff. There are two forces operating: distorting perceptions distorts beliefs – hence decisions – but it also distorts the technology that maps past outcomes into future probabilities of success. The first is bad, the second is good, and there is a tradeoff. Starting at the point where perceptions are correct ($\gamma = 0$), the distortion in decisions has a second order on welfare: the distortion applies to few realizations of costs, and for these realizations, the loss is small.[22] In contrast, the improvement in the technology has a first order effect on welfare. Formally,

$$\left.\frac{dV_\infty}{d\gamma}\right|_{\gamma=0} = \left.\frac{d\varphi_\infty}{d\gamma}\right|_{\gamma=0} \boldsymbol{\rho}'(\varphi^*) \Pr\{\varphi^* \geq c\}.$$

[22] We emphasize that the welfare reduction stemming from the distortion in decisions is a second order effect only when there is initially no distortion. There are clearly cases in which nonmarginal increases in confidence can be very costly.

Since $\left.\frac{d\varphi_\infty}{d\gamma}\right|_{\gamma=0} > 0$, we obtain:

Proposition: If $\boldsymbol{\rho}' > 0$, there exists a biased perception $\gamma > 0$ such that $V_\infty(\gamma) > V_\infty(0)$.

4.3. Robustness to alternative assumptions

The genesis of confidence. In our model, confidence depends on the agent's perception of the empirical frequency of past success. One motivation for this assumption is the idea that perceived bad average past performance would trigger anxiety, hence a decrease in performance. There are plausible alternative assumptions concerning the genesis of confidence (or lack of confidence), or more generally, how past perceptions affect current performance.

One plausible alternative assumption is that upon undertaking the project, the agent recalls a (possibly small) number of past experiences, and that recollections of failures among these past experiences decrease confidence or performance, for example because they generate unproductive intrusive thoughts. One could also imagine that recent past experience are more likely to be recalled.

Let us give two more specific examples:

(i) suppose that upon undertaking the project, the agent recalls one experience among all recorded past experiences, and that confidence κ is equal to 1 if that experience was a success, and equal to $1-\Delta$ (with $\Delta \in (0,1)$) if that experience was a failure. Under that specification, confidence is thus a random variable that may take two values (1 or $1-\Delta$), with the higher value having probability $\varphi = s/s+f$.

(ii) suppose that confidence depends on (and is increasing in) the fraction of successes obtained in the last T recorded experiences. Under that specification, confidence is a random variable that may take $T+1$ values.

The first example falls within the scope of our model, because it yields an expected performance equal to $\kappa(\varphi)\rho_0$, where $\kappa(\varphi) = 1-\Delta+\Delta\varphi$. We provide in the appendix an alternative form of biased perception that encompasses the second example where current performance depends on the particular history of past perceptions rather than just the summary statistic φ. A feature of that model is that a change in date t perceived outcome from failure to success can only increase later expected confidence. We show that our result holds with this alternative specification.

The welfare criterion. As discussed above, we evaluate the agent's welfare using long-run or steady state payoffs only. One alternative assumption is to evaluate welfare using discounted payoffs. With a discount factor sufficiently close to one, our insight would clearly remain valid.

For smaller discount factors, we would have to analyze how biased perceptions affect welfare in the transition to steady state. In transition however, our argument does not apply. Since beliefs may not be correct in transition, the distortion in beliefs induced by biased perceptions may not be second order: depending on whether the agent is currently too optimistic or too pessimistic, the bias may either increase or reduce the period during which beliefs remain far away from the truth. Thus we cannot be assured that the improvement in technology dominates the distortion of decisions; as a result the effect of the bias is indeterminate.

Nevertheless, if one were to evaluate welfare using discounted payoffs, it might be legitimate to allow for a bias in perception γ that would be time dependent. In that case, our model suggests that welfare is increased when γ is *positive in the long-run*.

5. Discussion

We have identified circumstances in which having biased perceptions increases welfare. How should one interpret this result? Our view is that it is reasonable to expect that in such circumstances agents will end up holding optimistic beliefs; for these environments, the welfare that they derive when they are optimistic is greater than when their beliefs are correct.

However, we do not conceive optimism as being the outcome of a deliberate choice by the agent. In our model, agents do not choose perceptions or beliefs. Rather, these are the product of recalled past experiences, which are themselves based on the types of attributions agents make, upon failing for example. Our view is that there are undoubtedly limits to the extent to which perceptions can be biased, but that it is reasonable to expect that the agents' attributions adjust *some* so as to generate some bias in perceptions.

We have described earlier a particularly simple form of biased attribution. We do not, however, suggest that precisely this type of attributions should arise. Other types of attributions as well as other types of information processing biases may lead to optimistic beliefs.

We start by suggesting below some examples. We will then question the main predictions of our model: that *rational agents* would persistently have biased perceptions, and be persistently overconfident about their chances of success.

Alternative information processing biases. In the model analyzed above, the agent sometimes attributes the cause of failure to events that are not likely to occur again in the future, while he never makes such attributions in case of success (non-permanence of failure/permanence of success). An alternative way in which an agent's perceptions may be biased is that rather than attributing failures to transient circumstances, he attributes successes to a broader set of circumstances than is appropriate;

in the psychology literature, this is referred to as pervasiveness of success bias (see e.g., Seligman 1990).

To illustrate this idea of pervasiveness, consider again the lawyer example. Suppose that cases that are available to the lawyer can be of two types: high profile cases that will attract much popular attention and low profile cases that will not. Suppose that in high profile cases, there will be many more people in court, including reporters and photographers, and consequently, much more stress, which we assume impairs performance. Suppose that one day the lawyer wins a case when there was no attention. Although the lawyer realizes that success would have been less likely had there been attention, he thinks "The arguments I made were so good, I would have won even if this had been a high profile case." In thinking this way, the lawyer is effectively recording the experience as a success whether it had been high profile or not.[23]

Another channel through which beliefs about chances of success may become biased is that the agent has a biased memory technology, and, for example, tends to remember more easily past successes than past failures.[24] Optimism may also stem from *self-serving bias*. Such bias occurs in settings where there is not a compelling, unique way to measure success. When one gives a lecture, he might naturally feel it was a success if the audience enthusiastically applauds when he's finished. If, unfortunately, that doesn't occur, but an individual comes up to me following the lecture and tells him that it was very stimulating, he can choose that as a signal that the lecture was a "success".

Can a rational agent fool himself? One prediction of our model is that when perceptions affect performance, rational agents prefer that their perceptions are biased. However, can a rational agent *fool* himself?

Imagine, for example, an agent who sets his watch ahead, in an attempt to ensure that he will be on time for meetings. Suppose that when he looks at his watch, the agent takes the data provided by the watch at face value. In this way, the agent fools himself about the correct time, and he easily arrives in time for each meeting. One suspects, however, that this trick may work once or twice but that eventually the agent will become aware of his attempt to fool himself, and not take the watch time at face value.

[23] More generally, this is an example of how inferences made at the time of the activity affect later recollection of past experiences. This inability to distinguish between outcomes and inferences when one attempts to recall past events constitutes a well-documented source of bias. A compelling example demonstrating this is given by Hannigan and Reinitz (2001). Subjects were presented a sequence of slides depicting ordinary routines, and later received a recognition test. The sequence of slides would sometimes contain an effect scene (oranges on the floor) and not a cause scene (woman taking oranges from the bottom of pile), and sometimes a cause scene and not an effect scene. In the recognition test, participants mistook new cause scenes as old when they had previously viewed the effect.

[24] See Morris (1999) for a survey of the evidence that an individual's mood affects his or her recollection process, e.g.

There are two important differences between this example and the problems we address. First, in the example the agent actively makes the decision to "fool" himself, and sets his watch accordingly. In our model, one should not think of the agent *choosing* to bias his perceptions, but rather the bias in perceptions is subconscious, and arises to the extent that the bias has instrumental value, that is, increases the welfare of people with the bias. The second difference is that the individual who sets his watch ahead of time will get constant feedback about the systematic difference between correct time and the time shown on his watch. Our model is aimed at problems in which this immediate and sharp feedback is absent. An academic giving a talk, a student writing a term paper for a particular class and a quarterback throwing a pass have substantial latitude in how they perceive a quiet reception to their talk, a poor grade on the paper or an incompletion. Each can decide that his performance was actually reasonably good and the seeming lack of success was due to the fault of others or to exogenous circumstances. Importantly, an individual with biased perceptions may never confront the gap in the way one who sets his watch ahead would. While it may be hard for people actively to fool themselves, it may be relatively easy (and potentially beneficial) to be passively fooled.[25]

How easy is it to be fooled? This presumably depends on the accuracy of the feedback the agent gets about his own performance. One can expect that fooling oneself is easier when the outcome is more ambiguous, because it is then easier to interpret the outcome as a success. Along these lines, putting agents in teams is likely to facilitate biased interpretations of the outcomes, and make it easier for each agent to keep one's confidence high: one can always put the blame on others in case of failures.[26] These observations suggest that when confidence affects performance, there may be a value to designing activities in a way that facilitates biases in information processing.

What if the agent is aware that confidence matters? A related objection to our model might be that when forming beliefs, a more sophisticated agent would take into account the fact that the data may be biased. This would be a reasonable assumption if the agent is aware that confidence matters and understands the possible effect that this may have on his attributions. Would we still obtain optimistic beliefs (hence overconfidence) as a prediction of our model if we allowed the agent to re-assess his beliefs over chances of success, to account for a possible over-representation of successes in the data?

The bias in perception clearly has instrumental value, but the resulting bias in belief (hence the agent's overconfidence) has no instrumental value. The agent might try

[25] This is, of course, not an argument that there are no limits on the extent to which an agent can be fooled. We only argue that it is plausible that the limit is not zero.

[26] There may be, of course, a cost to agents in getting less reliable feedback that offsets, at least partially, the gain from the induced high confidence.

reassessing his beliefs over chances of success so as to benefit from biased perceptions (through increased performance), without incurring the cost of suboptimal decision making induced by overconfidence.

However, these second thoughts in evaluating chances of success could limit the benefit that stems from biased perception. Reevaluating his chances of success may lead the agent to reassess his perceptions over past successes and failures as well. Introspection about his beliefs may lead him to recall past failures, otherwise forgotten, leading in turn to diminished future performance. Thus, even if the agent could learn to reevaluate his chance of success in a way that maximizes long run welfare, we should expect a reduced reevaluation that does not completely offset the bias in perception.

More formally, sophisticated agents who understand that their data may be biased, might adjust their prediction of success on the next try down, to say $p = \mu\beta(s, f)$, where $\mu \leq 1$ captures the extent of the adjustment. (The stronger the bias – as perceived by the agent – the greater the adjustment.) As a result, the perception φ would differ from the belief p in the long run, and the agent's overconfidence might be reduced. The comments above, however, suggest that for these agents, performance would depend not only on first hand perception φ, but also on the extent to which the agent thinks his data is biased:

$$\kappa = \kappa(\varphi, \mu),$$

where κ is increasing in μ (a larger adjustment diminishes performance).[27]

We have examined the case where μ is set equal to 1, that is, the agent is unaware that his perceptions are biased and there is no adjustment. We note that for any fixed $\mu < 1$ our analysis would carry over to this more general setting. For any fixed $\mu < 1$, there exists a bias γ^* that exactly offsets the adjustment μ, and leads to correct beliefs in the long run. From $\gamma^* = \gamma^*(\mu)$, a small further increase in the bias must be welfare increasing because the induced distortion in decisions has only a second order effect on welfare. In other words, if a sophisticated agent moderates his prediction of success on the next attempt to offset the bias in perception, a further bias in perception is welfare increasing.

But in this more general setting, our argument applies even if the bias in perception γ is fixed, or if no further bias is possible. For any fixed bias γ, there exists an adjustment μ that exactly offsets the bias. From $\mu^*(\gamma)$, a small increase in μ (that is, less adjustment of the agent's beliefs) is welfare increasing as well. Thus if both the bias γ and the adjustment μ that the agent thinks appropriate tend to evolve in a way that maximizes long-run welfare, then we should expect that the agent under-estimates the extent to

[27] Whether an adjustment has a large or small effect on performance could, for example, depend on whether the agent can simultaneously hold two somewhat contradictory beliefs – biased perceptions and unbiased, or less biased beliefs. This ability is clearly welfare enhancing.

which the data is biased (hence the overconfidence).[28] Hence, while in our main model, the decision maker was unaware of the possibility that his data was biased (i.e., $\gamma > 0$ but $\mu = 1$), in this extended model, the agent will be overconfident if he is simply unaware of the *extent* to which the data is biased ($\gamma > \gamma^*(\mu)$), a much weaker assumption.

Finally, we wish to point out a simple extension of our model that would generate overconfidence, even if no relationship between p and φ is imposed. Though we have assumed that confidence depends solely on perceptions, a quite natural assumption would be that confidence also depends on the beliefs of the agent about the likelihood of success, that is,

$$\kappa = \kappa(p, \varphi)$$

where both beliefs and perceptions may have a positive effect on confidence. This is a simple extension of our model that would confer instrumental value to beliefs being biased.[29]

5.1. Related literature.

There is a growing body of literature in economics dealing with overconfidence or confidence management. We now discuss several of these papers and the relation to our work.

One important modelling aspect concerns the treatment of information. In Benabou and Tirole (2002) and Koszegi (2000), the agent tries to assess how able he is, in a purely Bayesian way: He may actively decide to selectively forget some negative experiences (Benabou Tirole), or stop recording any further signal (Koszegi), but he is perfectly aware of the way information is processed, and properly takes it into account when forming beliefs. Thus beliefs are not biased on average; only the distribution over posterior beliefs is affected (which is why referring to confidence management rather than overconfidence may be more appropriate to describe this work). Another notable difference is our focus on long-run beliefs. For agents who in the long run have a precise estimate of their ability, the Bayesian assumption implies that this estimate must be the correct one (with probability close to 1). So in this context, confidence management essentially has only *temporary* effects.

Concerning the treatment of information, our approach is closer in spirit to that of Rabin and Schrag (1999). They analyze a model of "confirmatory bias" in which agents

[28] Depending on the shape of κ, more sophistication (i.e. $\gamma < 1$) need not be welfare increasing.

[29] More generally, our insight could be applied to psychological games (see, e.g., Geanakoplos, Pearce and Stacchetti (1989), and Rabin (1993)). In these games, beliefs affect utilities, so if agents have some (possibly limited) control over their attributional styles, they should learn to bias their beliefs in a way that increases welfare, and in the long-run, we should not expect beliefs to be correct. Our analysis thus questions whether the standard assumption that beliefs should be correct in equilibrium is a valid one.

form first impressions and bias future observations according to these first impressions. In forming beliefs, agents take their "observations" at face value, without realizing that their first impressions biased what they believed were their "observations". Rabin and Schrag then focus on how this bias affects beliefs in the long-run, and find that on average the agent is overconfident, that is, on average, he puts more weight than a Bayesian observer would on the alternative he deems most likely. A notable difference with our work is that in their case the bias has no instrumental value.

Another important aspect concerns the underlying reason why information processing biases may be welfare enhancing. In Waldman (1994) or Benabou and Tirole (2002), the benefit stems from the fact that the agent's criterion for decision making does not coincide with welfare maximization, while in Koszegi, the benefit stems from beliefs being directly part of the utility function.[30]

Waldman (1994) analyzes an evolutionary model in which there may be a divergence between private and social objectives. He considers a model of sexual inheritance of traits (no disutility of effort/disutility of effort; correct assessment of ability/overconfidence), in which fitness would be maximized with no disutility of effort and correct assessment of ability. Disutility of effort leads to a divergence between the social objective (offspring) and the private objective (utility). As a result, individuals who possess the trait "disutility of effort" have higher fitness if they are overconfident (because this induces greater effort).[31]

In Benabou and Tirole (2002), the criterion for decision making also diverges from (ex ante) welfare maximization. Benabou and Tirole study a two-period decision problem in which the agent has time-inconsistent preferences. As a consequence of this time inconsistency, the decision criterion in the second period does not coincide with ex ante welfare maximization, and biased beliefs in this context may induce decisions that are better aligned with welfare maximization..

To illustrate formally why a divergence between private objectives and welfare gives instrumental value to biased beliefs, assume that in our model, performance is independent of beliefs, but that the agent only undertakes the activity when

$$p > \beta c \text{ with } \beta > 1. \tag{5.1}$$

[30] See also Samuelson (2001) for another example of a model in which an incorrectly specified decision criterion leads to welfare enhancing biases, and Fang (2001) and Heifetz and Spiegel (2001) for examples of models where, in a strategic context, beliefs are assumed to be observable, hence directly affect opponent's reactions.

[31] Waldman (1994)'s main insight is actually stronger: he shows that the trait "disutility of effort", and hence the divergence between private and social objectives may actually persist: the combination of disutility of effort and overconfidence in ability can be evolutionary stable. The reason is that for individuals who possess the trait "overconfidence" disutility of effort is optimal, so the combination disutility of effort and overconfidence in ability is a local maximum for fitness (only one trait may change at a time).

One interpretation is that the agent lacks the will to undertake the risky activity, and that he only undertakes it when its expect return exceeds one by a sufficient amount.[32]

Given this criterion for making the decision whether to undertake the risky activity or not, we may write expected welfare as a function of the bias γ. We obtain:

$$w(\lambda) = \int_{\psi^\gamma(\rho)>\beta c} (\rho - c)g(c)dc,$$

which implies

$$w'(\gamma) = (\rho - \frac{\psi^\gamma(\rho)}{\beta})\frac{d\psi^\gamma(\rho)}{d\gamma} g(\frac{\psi^\gamma(\rho)}{\beta}).$$

Hence $w'(\gamma)$ has the same sign as $\rho - \frac{\psi^\gamma(\rho)}{\beta}$, and is positive at $\gamma = 0$ (correct beliefs), since $\psi^0(\rho) = \rho$ and $\beta > 1$. Intuitively, biased beliefs allow the agent to make decisions that are better aligned with welfare maximization.

Finally, Van den Steen (2002) provides another interesting model of overconfidence, that explains why most agents would in general believe they have higher driving ability than the average population. Unlike some of the work mentioned above, this overconfidence does not stem from biases in information processing but from the fact that agents do not use the same criterion to evaluate their strength..

[32] This formulation can actually be viewed as a reduced form of the agent's decision problem in Benabou Tirole (2001), where β is interpreted as a salience for the present. To see why, assume that (i) there are three periods, (ii) the decision to undertake the activity is made in period 2; (iii) the benefits are enjoyed in period 3; (iv) there is no discounting. The salience for the present has the effect of inflating the cost of undertaking the activity in period 2, hence induces a divergence from ex ante welfare maximization (that would prescribe undertaking the activity whenever $p > c$).

Bibliography

Aronson, J., M. Lustina, C. Good, and K. Keough (1999), "When White Men Can't Do Math: Necessary and Sufficient Factors in Stereotype Threat," *Journal of Experimental Social Psychology* 35, 29-46.

Baker, L, and R. Emery (1993), "When Every Relationship is Above Average: Perceptions and Expectations of Divorce at the Time of Marriage," *Law and Human Behavior*.

Baker, S., C. Firth, and R. Dolan (1997), "The Interaction between Mood and Cognitive Function Studied with PET," *Psychological Medicine*, 27, 565-578.

Benabou, R. and J. Tirole (2002), "Self-Confidence and Personal Motivation", *Quarterly Journal of Economics*, 117, 871-915.

Brand, N., L. Verspui and A. Oving (1997), "Induced Mood and Selective Attention," *Perceptual and Motor Skills*, 84, 455-463.

Ellis, H. S. Ottaway, L. Varner, A. Becker and B. Moore (1997), "Emotion, Motivation, and Text Comprehension: The Detection of Contradictions in Passages," *Journal of Experimental Psychology: General*, 126(2), 131-46.

Fang, H. (2001), "Confidence Management and the Free-Riding Problem," mimeo, Yale University.

Fischoff, B., Slovic, P., and Lichtenstein, S. (1977), "Knowing with Certainty: The Appropriateness of Extreme Confidence," *Journal of Experimental Psychology: Human Perception and Performance,* 3(4), 552-564.

Geanakoplos, J., D. Pearce and E. Stacchetti (1989), "Psychological Games and Sequential Rationality," *Games and Economic Behavior* 1, 60-79.

Guthrie, C., J. Rachlinski and A. Wistrich (2001), "Inside the Judicial Mind: Heuristics and Biases," *Cornell Law Review* 86, 777-830.

Hannigan, S. L. and M. T. Reinitz (2001), "A Demonstration and Comparison of Two Types of Inference-Based Memory Errors", *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 27 (4) 931-940.

Heifetz, A. and Y. Spiegel (2001), "The Evolution of Biased Perceptions", mimeo.

Koszegi, B. (2000), "Ego Utility, Overconfidence and Task Choice," mimeo.

Lichtenstein, S., B. Fischoff, and L. Phillips (1982), "Calibration of Probabilities: The State of the Art to 1980," in D. Kahneman, P. Slovic, and A. Tversky, (eds.), *Judgement under Uncertainty: Heuristics and Biases*, pp. 306-334; Cambridge: Cambridge University Press.

McKenna, F. and C. Lewis (1994), "A Speech Rate Measure of Laboratory Induced Affect: The Role of Demand Characteristics Revisited," *British Journal of Clinical Psychology,* 33, 345-51.

Morris, W. N. (1999), "The Mood System," in *Well-Being: The Foundations of Hedonic Psychology*, D. Kahneman, E. Diener and N. Schwartz, eds., New York, NY: Russell Sage Foundation.

Nisbett, R. and L. Ross (1980), *Human Inference: Strategies and Shortcomings of Social Judgment,* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Rabin, M. (1993), "Incorporating Fairness into Game Theory and Economics", *American Economic Review*, 83, 1281–1302

Rabin, M. and J. Schrag (1999), "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics,* 1, pp. 37-2.

Ross, M. and F. Sicoly (1979), "Egocentric Biases in Availability and Attribution," *Journal of Personality and Social Psychology* 322.

Samuelson, L. (2001), "Information-Based Relative Consumption Effects," mimeo.

Seligman, M. (1990), *Learned Optimism,* New York: Knopf.

Steele, C. M., & Aronson, J. (1995), "Stereotype Threat and the Intellectual Test Performance of African-Americans," *Journal of Personality and Social Psychology*, 69, 797-811.

Steele, C. M., & Aronson, J. (1998), "How Stereotypes Influence the Standardized Test Performance of Talented African American Students," in C. Jencks & M. Phillips (Eds.), *The Black-White Test Score Gap.* Washington, D. C.: Brookings Institution, 401-427.

Svenson, O. (1981), "Are We All Less Risky and More Skillful Than Our Fellow Drivers?", *Acta Psychologica* 143, 145-46.

Taylor, S. and J. Brown (1988), "Illusion and Well-Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin* 103, 193-210.

Teasdale, J. and M. Russell (1983), "Differential Effects of Induced Mood on the Recall of Positive, Negative and Neutral Moods," *British Journal of Clinical Psychology*, 22, 163-171.

Tversky, A. and D. Kahneman (1973), "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Science*, 5, 207-232.

Van den Steen, E. (2002), "Skill or Luck? Biases of Rational Agents," mimeo.

Waldman, M. (1994), "Systematic Errors and the Theory of Natural Selection," *American Economic Review* 84, 482-497.

Weinstein, N. (1980), "Unrealistic Optimism about Future Life Events," *Journal of Personality and Social Psychology* 39, 806-820.

Appendix A. Convergence result.
Recall that $\rho(\varphi) = \rho_0 \kappa(\varphi)$. Define $\widetilde{\rho} = \psi^\gamma \circ \rho$, that is:

$$\widetilde{\rho}(\varphi) = \frac{\rho(\varphi)}{\rho(\varphi) + (1 - \rho(\varphi))(1 - \gamma)}$$

We first assume that the equation:

$$\varphi = \widetilde{\rho}(\varphi) \tag{5.2}$$

has a unique solution, denoted φ^*. We wish to show that in the long run the perception φ gets concentrated around φ^*. Formally, we will show that for any $\varepsilon > 0$,[33]

$$\underline{\lim}_t \Pr\{\varphi_t \in [\varphi^* - \varepsilon, \varphi^* + \varepsilon] = 1.$$

Our proof builds on a large deviation result in statistics. The proof of this result, as well as that of the Corollary that builds on it are relegated to Appendix B.

Lemma 1 *(large deviations) Let $X_1, ..., X_n$ be n independent random variable, and let $S = \sum_i X_i$. Assume $\mid X_i \mid < M$ for all i. Define $\bar{m} = \frac{1}{n} E \sum_i X_i$ and $h(x, S) = \sup_{t>0} xt - \frac{1}{n} \sum_i \ln E \exp tX_i$. For any $x > \bar{m}$, we have:*

$$\Pr\{\frac{1}{n} \sum_i X_i > x\} \leq \exp -nh(x, S)$$

Besides, for any $x > \bar{m}$, $h(x, S) > h_0$ for some $h_0 > 0$ that can be chosen independently of n.

To assess how the perception φ evolves from date t to date $t + T$, we derive below bounds on the number of outcomes N and the number of successes S that are recorded during that period of time. These bounds will be obtained as a Corollary of Lemma 1.

Formally, consider the random variable $\widetilde{y}_t$ that takes value 0 or 1 depending on whether the outcome is failure or a success, and the random variable $\widetilde{z}_t$ that takes value 1 or 0 depending on whether the agent is (or is not) subject to a perception bias at t (the outcome is not recorded in the event $y_t = 0, z_t = 1$). Now consider a sequence of T projects undertaken at date $t, ..., t+T-1$, and suppose that performance at any of these dates is bounded above by $\bar{\rho}$. The numbers S of recorded success and N of recorded outcomes are respectively $S = \sum_{i=0}^{T-1} \widetilde{y}_{t+i}$. and $N = \sum_{i=0}^{T-1} \widetilde{y}_{t+i} + (1 - \widetilde{y}_{t+i})(1 - \widetilde{z}_{t+i})$. We have:

[33] In what follows, date t corresponds to the t^{th} attempt to undertake the project. That is, we ignore dates at which the project is not undertaken. This can be done without loss of generality because at dates where the project is not undertaken, the perception ϕ does not change.

Corollary 2 *Assume $\rho_{t+i} \leq \bar{\rho}$ for all $i \in \{0, .., T-1\}$. Then for any small $x > 0$, there exists $h > 0$ that can be chosen independently of T such that*

$$\Pr\{\frac{N}{T} > (1-\gamma)/2 \text{ and } \frac{S}{N} < \frac{\bar{\rho}}{\bar{\rho} + (1-\bar{\rho})(1-\gamma)} + x\} \geq 1 - \exp -hT$$

Corollary 2 implies that when perceptions remain below φ, so that performance remains bounded above by $\rho(\varphi)$, then the ratio of recorded successes S/N cannot exceed $\widetilde{\rho}(\varphi)$ by much. Convergence to φ^* will then result from the fact that as long as φ exceeds φ^*, $\widetilde{\rho}(\varphi)$ is below φ. More precisely, since $\widetilde{\rho}$ is smooth, since $\widetilde{\rho}(0) > 0$ and $\widetilde{\rho}(1) < 1$, and since (5.2) has a unique solution, we have:

Lemma 3 *For any $\varepsilon > 0$, there exists $\alpha > 0$ such that for any $\varphi > \varphi^* + \varepsilon$, $\widetilde{\rho}(\varphi) < \varphi - \alpha$.*

In words, when current perceptions exceed φ^* by more than ε, the adjusted probability of success $\widetilde{\rho}$ is below current perceptions by more than α. Combining this with Corollary 2, we should expect perceptions to go down after a large enough number of periods. The next Lemma makes this statement precise.

Lemma 4 *Fix $\varepsilon > 0$ and $\alpha \in (0, \varepsilon)$ as in Lemma 3, and let $\beta = \frac{\alpha}{2}\frac{1-\gamma}{2}$. There exists $h > 0$ and $\underline{t} > 0$ such that, for any date $t_0 \geq \underline{t}$, for any $T \leq \beta t_0$, and for any $\varphi \geq \varphi^* + \varepsilon$,*

$$\Pr\{\varphi_{t_0+T} < \varphi_{t_0} - \frac{T}{t_0 + T}\frac{\beta}{2} \mid \varphi_{t_0} = \varphi\} \geq 1 - \exp -hT$$

P roof. Let $n_{t_0} = s_{t_0} + f_{t_0}$. For $\underline{t}$ large enough and $t_0 \geq \underline{t}$, the event $n_{t_0} \leq (1-\gamma)t_0/2$ has probability at most equal to $\exp -ht_0$ for some h independent of t_0. Consider now the event $n_{t_0} > (1-\gamma)t_0/2$. Let S and N be the number of successes and recorded events during $\{t_0, ..., t_0 + T - 1\}$. We have:

$$\varphi_{t_0+T} = \frac{s_{t_0} + S}{n_{t_0} + N} = \frac{n_{t_0}}{n_{t_0} + N}\varphi_{t_0} + \frac{N}{n_{t_0} + N}\frac{S}{N} = \varphi_{t_0} + \frac{N}{n_{t_0} + N}(\frac{S}{N} - \varphi_{t_0}) \qquad (5.3)$$

During this period of time, φ_t remains below $\varphi_{t_0} + T/n_0$, hence below $\varphi_{t_0} + \alpha/2$. It follows that ρ_t remains below $\rho(\varphi_{t_0} + \alpha/2)$. We now apply Corollary 2 with $x = \alpha/4$, and choose h accordingly. The event $\frac{S}{N} < \widetilde{\rho}(\varphi_{t_0} + \alpha/2) + \alpha/4$ and $N > (1-\gamma)T/2$ thus has probability $1 - \exp -hT$. Under that event (and using Lemma 3)), we have $\frac{S}{N} - \varphi_{t_0} < -\alpha/4$, which implies (using 5.3):

$$\varphi_{t_0+T} < \varphi_{t_0} - \frac{(1-\gamma)T/2}{t_0 + (1-\gamma)T/2}\alpha/4. \blacksquare$$

We can now conclude the proof. We will show that for t_0 large enough (i) for any $\varphi_{t_0} > \varphi^* + \varepsilon$, the perception φ eventually gets below $\varphi^* + \varepsilon$ with probability close to 1, and (ii) if $\varphi_{t_0} < \varphi^* + \varepsilon$, then the perception φ *always* remains below $\varphi^* + 2\varepsilon$ with probability close to 1.

(i) This is obtained by iterative application of Lemma 4. We set $t_k = (1+\beta)t_{k-1}$. So long as $\varphi_{t_k} > \varphi^* + \varepsilon$, we may apply Lemma 4 and get that $\varphi_{t_{k+1}} < \varphi_{t_k} - \mu$ with probability $1 - \exp -h\beta t_k$, where $\mu = \beta^2/(2(\beta+1))$. We thus obtain (after at most $K = (1-\varphi^*)/\mu$ iterations):

$$\Pr\{\exists t < t_K,\, \varphi_t < \varphi^* + \varepsilon \mid \varphi_{t_0} \geq \varphi^*\} \geq (1 - \exp -h\beta t_0)^K. \tag{5.4}$$

Since K and h are independent of t_0, the right hand side is arbitrarily close to 1 when t_0 is large.

(ii) Consider the event A_s where φ gets above $\varphi^* + 2\varepsilon$ for the first time (after t_0) at date s. For any $s > t_0$, choose t_s such that $(1+\beta)t_s = s$. Under the event $n_t > (1-\gamma)t/2$, at least $(1-\gamma)\varepsilon t/2 (\geq \beta t)$ periods are required for φ to get from $\varphi^* + \varepsilon$ to $\varphi^* + 2\varepsilon$, so we must have $\varphi_{t_s} \geq \varphi^* + \varepsilon$ and we may apply Lemma 4 to obtain

$$\Pr\{A_s \mid n_{t_s} > (1-\gamma)t_s/2\} \leq \exp -h_0 s,$$

for some h_0 independent of s, which further implies:

$$\Pr\{\bigcup_{s>t_0} A_s\} \leq \sum_{s>t_0} \Pr\{n_{t_s} < (1-\gamma)t_s/2\} + \Pr\{A_s \mid n > (1-\gamma)t_s/2\} \leq \exp -ht_0 \tag{5.5}$$

for some h independent of t_0.

Combining (5.4) and (5.5), we finally obtain

$$\underline{\lim}\, \Pr\{\varphi_t < \varphi^* + \varepsilon\} = 1.$$

Proving that for any $\varepsilon > 0$, $\underline{\lim}\, \Pr\{\varphi_t > \varphi^* - \varepsilon\} = 1$ can be done similarly.

Generalization.

Convergence: Our convergence result can be easily generalized to the case where equation (5.2) has a finite number of solutions. To fix ideas, consider the case where there are three solutions denoted, φ_i^*, $i = 1,2,3$, with $\varphi_1^* < \varphi_2^* < \varphi_3^*$. While Lemma 3 no longer holds, the following weaker statement does:

> For any $\varepsilon > 0$, there exists $\alpha > 0$ such that for any $\varphi \in (\varphi_1^* + \varepsilon, \varphi_2^* - \varepsilon)$, and any $\varphi \geq \varphi_3^* + \varepsilon$, $\widetilde{\rho}(\varphi) < \varphi - \alpha$.

Lemma 4 thus applies to any $\varphi \in (\varphi_1^* + \varepsilon, \varphi_2^* - 2\varepsilon)$ or any $\varphi \geq \varphi_3^* + \varepsilon$, and convergence results as before from (iterative) application of Lemma 4. For example, if $\varphi_{t_0} \leq \varphi_2^* - 2\varepsilon$, then with probability $1 - \exp -ht_0$, φ_t will get below $\varphi_1^* + \varepsilon$ and never get back above $\varphi_1^* + 2\varepsilon$.

Monotonicity: To prove our main result (Proposition 1), we used the fact that when γ increases, the (unique) solution to equation (5.2) increases as well. When there are several solutions to (5.2), this property does not hold for all solutions. In particular, it does not hold when for the solution φ_i^* considered, $\widetilde{\rho}'(\varphi_i^*) > 1$, as would typically be the case for $i = 2$. Nevertheless, these solutions are unstable: even if convergence to such unstable solutions occurs with positive probability for some values of γ,[34] a slightly higher value of γ would make these trajectories converge to a higher (and stable) solution.

Intuitively, a higher bias has two effects: (i) it reduces the number of failures that are recorded (failures that would have been recorded under γ are not recorded with probability $\gamma' - \gamma$), hence it mechanically increases the perception of past success rate φ_t (ii) Because φ_t increases, there is also an indirect effect through increased confidence: successes are more likely.

Both these effects go in the direction of increasing φ_t, and the first effect alone increases φ_t by a positive factor with probability close to 1. When $\varphi_2^*(\gamma') \leq \varphi_2^*(\gamma)$, trajectories leading to $\varphi_2^*(\gamma)$ when the bias is equal to γ must lead (with probability one) to trajectories leading to higher long-run perception when the bias is equal to γ'. The only possible candidate is then $\varphi_3^*(\gamma')$.

Robustness.

We next examine the robustness of our result to alternative formulations where performance at date t depends on the sequence of the most recent perceived outcomes, rather than on the summary statistic φ.

Formally, at any date t at which the project is undertaken, the outcome y_t may take two values, and we set $y_t = 0$ or 1 depending on whether the outcome is failure or a success. We consider the possibility that the agent's perception is biased as follows: with probability γ, a failure is actually perceived as a success. We denote by Y_t the history of outcomes, by $\widehat{y}_t$ the perceived outcome at t, and by $\widehat{Y}_t$ (respectively $\widehat{Y}_t^T$) the history of perceived outcomes (respectively last T perceived outcomes).[35] We assume that performance at t depends on the sequence $\widehat{Y}_t^T$,

$$\rho_t = \rho(\widehat{Y}_t^T),$$

and that performance at t is increasing in $\widehat{Y}_t^T$, that is, there exists a positive number a

[34] Convergence to such unstable solutions cannot be excluded for some values of γ.

[35] Note that we omit dates at which the project is not undertaken, because performance will be assumed to depend only on the sequence of perceived outcomes, and not on the actual date at which these outcomes occured. Thus, we have $Y_t = (y_1, ..., y_{t-1})$, $\widehat{Y}_t = (\widehat{y}_1, ..., \widehat{y}_{t-1})$ $\widehat{Y}_t^T = \widehat{Y}_t$ for $t \leq T$, and $\widehat{Y}_t^T = (\widehat{y}_{t-T}, ..., \widehat{y}_{t-1})$ for $t > T$.

such that for all histories $\widehat{Y}^{T-1}$ (of length $T-1$),

$$\rho(Y^{T-1}, 1) - \rho(Y^T, 0) \geq a.$$

As in our model, the triplet (ρ, β, γ) induces a probability distribution over beliefs and histories of perceptions. Although performance no longer converges to a point distribution, it is standard to show that $\widehat{Y}_t^T$ converges to a unique invariant distribution. It is not difficult then to show that the perceived frequency of success, hence beliefs, also converges to some $p(\gamma)$.[36] Long-run expected welfare may thus be written as:

$$V(\gamma) = E_\gamma \int_{c \leq p(\gamma)} [\rho(\widehat{Y}_t^T) - c] g(c) dc.$$

As in our model, it then follows that starting from $\gamma = 0$ (where in the long run beliefs are correct and coincide with $\lim E_0 \rho(\widehat{Y}_t^T)$), a positive bias $\gamma > 0$ induces a distortion in decisions that has a second order effect on welfare. To show that our result carries over to this setting, we need to check that the increase in γ generates a first order increase in long run expected performance.

To verify this we consider the two stochastic processes associated respectively with $\gamma = 0$ and $\gamma = \gamma_0 > 0$. Consider a sequence $Z = (z_t)_{t \geq 1}$ where $z_t \in \{0, 1\}$ for all t, that captures whether the agent may or may not be subject to a perception bias at t. Formally, we define the function $\widehat{y}_t = b(y_t, z_t)$ such that $\widehat{y}_t = 1$ if $y_t = 1$ or $z_t = 1$, and $\widehat{y}_t = 0$ otherwise.[37] The function $\widehat{Y}_t = B_t(Y_t, Z_t)$, where Z_t denotes the truncated sequence $(z_1, ... z_{t-1})$ is defined accordingly. The first stochastic process ($\gamma = 0$) corresponds to the case where $z_t = 0$ for all t. The second stochastic process corresponds to the case where $(z_t)_{t \geq 1}$ is a sequence of i.i.d. random variable that takes value 0 with probability $1 - \gamma_0$ and 1 with probability γ_0.

Now denote by $\widetilde{X} = (\widetilde{x}_t)_{t \geq 1}$ a sequence of i.i.d. random variables uniformly distributed on $[0, 1]$. The outcome at date t can be viewed as a function of $\widehat{Y}_t$ and the realization x_t, where $y_t = R_t(\widehat{Y}_t, x_t) = 1$ if $x_t < \rho(\widehat{Y}_t)$, and $y_t = 0$ otherwise. So the outcome Y_{t+1} can be viewed as a function of Z_t and X_{t+1}, say $Y_{t+1} = S_{t+1}(Z_t, X_{t+1})$. Since B_t is non-decreasing in Y_t and in Z_t, and since R_t is non-decreasing in $\widehat{Y}_t$, we obtain (by induction on t) that for all t, S_t is non-decreasing in Z_t.

For any given realization X of $\widetilde{X}$, let Y_t^0 denote the sequence of outcomes that obtains up to date t under the first stochastic process: $Y_t^0 \equiv S_t(0, X_t)$. For any given

[36] Alternatively, we could have assumed that the agent attempts to estimate the probability of success at the next trial, contingent on the last T perceived outcomes (and that these conditional beliefs each get close to the corresponding empirical frequencies). Conditional on $\widehat{Y}^T$, the agent's belief would then converge to some $p(\gamma, \widehat{Y}^T)$ that coincides with $\rho(\widehat{Y}_t^T)$ when there is no bias ($\gamma = 0$).

[37] Note that $b(y_t, z_t) = \max(y_t, z_t)$, so $b(., .)$ is non-decreasing in each of its arguments.

realizations (X_t, Z_t), the difference between expected performance at t under the two stochastic processes can be written as:

$$\rho(B_t(S_t(Z_{t-1}, X_t), Z_t) - \rho(S_t(0, X_t))$$

and since S_t is non-decreasing in Z_t, this difference is at least equal to

$$\rho(B_t(Y_t^0, Z_t)) - \rho(Y_t^0)$$

Then a first order increase in performance would obtain in the long run, because at any date, the event $(y_t, z_t) = (0, 1)$ has positive probability.[38]

[38] Note that the argument above is quite general, and applies even when performance depends on the whole history, as was the case in the main body of the paper. Indeed, in that case, when the number of trials becomes large, the number of realizations for which $(y_t, z_t) = (0, 1)$ is (with positive probability) a positive (and non-vanishing) fraction of the number of trials. So $\rho(B_t(Y_t^0, Z_t)) - \rho(Y_t^0)$ is non-negative, and bounded away from 0 with positive probability.

Appendix B. Large deviations results

Proof of Lemma 1: For any $z > 0$, we have:

$$\Pr\left\{\frac{S}{n} \geq x\right\} = \Pr\left\{z(\frac{S}{n} - x) \geq 0\right\} = \Pr\{\exp z(\frac{S}{n} - x) \geq 1\} \leq E \exp z(\frac{S}{n} - x) \quad (5.6)$$

Since the random variables X_i are independent,

$$E \exp \frac{z}{n} \sum_i X_i = \prod_i E \exp \frac{z}{n} X_i.$$

We may thus write, letting $t = \frac{z}{n}$,

$$E \exp z(\frac{S}{n}) = \exp \ln \prod_i E \exp tX_i = \exp \sum_i \ln E \exp tX_i,$$

which implies, using **(5.6):**

$$\Pr\left\{\frac{S}{n} \geq x\right\} \leq \exp -n(tx - \frac{1}{n} \sum_i \ln \exp tX_i) \quad (5.7)$$

Since (5.7) holds for any $t > 0$, we obtain the desired inequality. Besides, since $\mid X_i \mid < M$ for all i, for t close to 0, $xt - \frac{1}{n}\sum_i \ln E \exp tX_i = t(x - \bar{m}) + O(t^2)$, hence for $x > \bar{m}$, $h(x, S)$ is bounded below by some $h_0 > 0$ that depends only on M and $x - \bar{m}$. ■

Proof of Corollary 2: For any small $x > 0$, we set x_0 such that

$$\frac{\bar{\rho} + x_0}{\bar{\rho} + x_0 + (1 - \bar{\rho} - x_0)(1 - \gamma - x_0)} = \frac{\bar{\rho}}{\bar{\rho} + (1 - \bar{\rho})(1 - \gamma)} + x.$$

Choose x small enough so that $x_0 < (1 - \gamma)/2$. Consider first the hypothetical case where performance would be equal to $\bar{\rho}$ at all dates. Then S is the sum of T independent random variables. Applying Lemma 1 to the sum S, we get that there exists $h_0 > 0$ that can be chosen independently of T such that:

$$\Pr_{\bar{\rho}}\{\frac{S}{T} \geq \bar{\rho} + x_0\} \leq \exp -h_0 T, \quad (5.8)$$

where the index $\bar{\rho}$ indicates that outcomes are generated assuming performance is equal to $\bar{\rho}$. Because performance actually depends on past realizations, S is not the sum of independent variables. However, under the assumption that performance remains bounded above by $\bar{\rho}$, we have:

$$\Pr\{\frac{S}{T} \geq \bar{\rho} + x_0\} \leq \Pr_{\bar{\rho}}\{\frac{S}{T} \geq \bar{\rho} + x_0\}$$

Consider now realizations ζ of S/T for which $\zeta \leq \bar{\rho}+x_0$. Conditional on a realization ζ of S/T, $\sum_i(1-z_{t+i})(1-y_{t+i})$ is the sum of $T-S=T(1-\zeta)$ independent and *identical* random variables, each having $(1-\gamma)$ as expected value. Applying Lemma 1, we thus obtain

$$\Pr\{\frac{N}{T} < \zeta + (1-\zeta)(1-\gamma-x_0) \mid \frac{S}{T} = \zeta\} \leq \exp -h_1(1-\zeta)n \tag{5.9}$$

for some $h_1 > 0$ that can be chosen independently of n. Set $h_2 = h_1(1-(\bar{\rho}+x_0))$. Since $\zeta + (1-\zeta)(1-\gamma-x_0) > (1-\gamma)/2$, and since for any $\zeta \leq \bar{\rho}+x_0$, (by definition of x_0)

$$\frac{\bar{\rho}}{\bar{\rho}+(1-\bar{\rho})(1-\gamma)} + x \geq \frac{\zeta}{\zeta+(1-\zeta)(1-\gamma-x_0)},$$

and inequality (5.9) implies

$$\Pr\{\frac{N}{T} < (1-\gamma)/2 \text{ or } \frac{S}{N} > \frac{\bar{\rho}}{\bar{\rho}+(1-\bar{\rho})(1-\gamma)} + x \mid \frac{S}{T} \leq \bar{\rho}+x_0\} \leq \exp -h_2 T$$

Combining this with (5.8), we obtain the desired inequality. ■