EGO UTILITY, OVERCONFIDENCE, AND TASK CHOICE

Botond Köszegi
University of California, Berkeley

Abstract
This paper models behavior when a decision maker cares about and manages her self-image. In addition to having preferences over material outcomes, the agent derives "ego utility" from positive views about her ability to do well in a skill-sensitive, "ambitious," task. Although she uses Bayes' rule to update beliefs, she tends to become overconfident regarding which task is appropriate for her. If tasks are equally informative about ability, her task choice is also overconfident. If the ambitious task is more informative about ability, she might initially display underconfidence in behavior, and, if she is disappointed by her performance, later become too ambitious. People with ego utility prefer to acquire free information in smaller pieces. Applications to employee motivation and other economic settings are discussed. (JEL: D83, D11)

1. Introduction

Most of the existing information economic models treat humans as scientists objectively studying reality. According to these models, we gather information in an unbiased manner, combine it with our beliefs in a detached, logical way, reach accurate inferences, and make generally good decisions. Psychological research, however, indicates that we instead resemble a scientist who cares more about feeling smart than about reality. According to this view, we use and interpret information in a large part to make favorable conclusions about ourselves, often damaging the quality of our decisions.

This paper studies such "self-image" aspects of information processing in a simple rational model in which beliefs enter the utility function directly: The agent derives "ego utility" from positive views about the self, holding constant standard utilitarian outcomes usually assumed relevant in economics. For example, a

Acknowledgments: I benefited from numerous enjoyable discussions with Mara Berman, Peter Diamond, Jonathan Gruber, Melissa Kearney, Bulcsú Köszegi, Chrys "Villy" Margaritides, Ted Miguel, Markus Möbius, Sendhil Mullainathan, Matthew Rabin, Tanya Rosenblat, and Chris Shannon. Two anonymous referees also contributed greatly to the paper. I am grateful for financial help from the Sloan Dissertation Fellowship and for the hospitality of the Central European University while some of this work was completed.

E-mail address: Köszegi: botond@econ.berkeley.edu

Journal of the European Economic Association June 2006 4(4):673–707

person might enjoy thinking that she is a good basketball player, even if that is unlikely to benefit her in any direct way. I examine how a decision maker's concern for her self-image affects and interacts with traditional, "instrumental" motivations for acquiring information about herself and choosing the right activities for her ability.

The decision maker faces a very simple problem. She can collect an arbitrary amount of free information about her ability to perform well in the more ambitious of two financial activities, one which only makes financial sense for higher types. She can also stop information collection at any time. She then decides whether to be ambitious or unambitious in each of two financial choice periods. Her utility depends on monetary outcomes as well as her beliefs about her ability, where the latter (ego utility) term is tightly linked to the financial choice problem she is facing: She likes to believe that she is "good enough" to engage in the ambitious activity. For methodological discipline—and without claiming that all ego-related phenomena are exclusively rational in their origin—I assume that the agent is limited in generating positive beliefs about herself by Bayes' rule. Even without relaxing standard assumptions on belief formation, the model predicts a host of behaviors traditional theories would have difficulty explaining in a unified setting, and has other, testable, implications.

It is important to note that the Bayesian framework (combined with the assumption of correct priors) implies that agents are always unbiased in their beliefs about the underlying ability parameter determining performance. Nevertheless, the first logical consequence of my setup is a systematic bias in the level of beliefs about the appropriate choice of activity: Agents become overconfident in their ability to perform the more ambitious task. That is, too many agents come to—rationally and honestly—believe that they would make more money on average with the ambitious option than the unambitious one. Intuitively, because agents want to believe that this is the case, as they voluntarily collect information they are more likely to stop when they believe it is.

In addition to affecting the level of beliefs, ego utility leads the agent to distort instrumental choices to manipulate the information she receives about herself.[1] If the decision maker is satisfied with her present beliefs, she might distort her choices to avoid receiving information about herself. I call this the *self-image protection* motive. On the other hand, if she is dissatisfied with her current perception of herself, she might go out of her way to try to improve her beliefs. This is the *self-image enhancement* motive. The model organizes how these two effects play themselves out across different categories of tasks and within a category of tasks over time. Although I discuss other determinants of behavior, the results focus on how the distortions caused by self-image protection

1. Throughout the paper, a "distortion" refers to suboptimal choices from the point of view of the instrumental side of utility only.

and enhancement depend on the time period and the relative informativeness of the ambitious and unambitious tasks.

If the ambitious and unambitious tasks are equally informative about ability, the decisionmaker cannot manipulate her beliefs through distorting task choice. Because she therefore does what she thinks is best financially based on her beliefs, the overconfidence in beliefs translates directly into overconfidence in actions.

If the ambitious task is more informative about ability, the decision maker's choices do not necessarily accord with what she herself thinks is financially optimal. On the one hand, when she thinks she is good enough to be ambitious, but she is not quite sure, she might avoid the ambitious option for fear of learning bad news about her ability. As a result, overconfidence in beliefs can be coupled with underconfidence in observed actions. In fact, among people who are identical other than in the importance they attach to ego utility, there is a negative relationship between confidence and participation in the ambitious task: Although average overconfidence increases in the weight attached to ego utility, the probability of choosing the ambitious option approaches zero as that weight approaches infinity.

On the other hand, if the agent tries the ambitious task in period 1 and she is so disappointed by her performance that she now thinks the unambitious option is financially more appropriate, the self-image enhancement motive might come into play. Hoping that a good financial outcome will improve her self-image, she chooses to be ambitious despite probable losses. Although protection of self-image can affect the decision maker at each stage of her participation in the task, self-image enhancement tends to happen only at later stages. Because she can costlessly learn about her ability before making financial choices, she would not begin by choosing the ambitious option just to learn about herself. In case she is disappointed in her early performance, however, she is willing to sacrifice financial payoffs to reassert her self-image.

Although all of the above results derive from the agent's attempts to manipulate the information she receives about herself, the model so far abstracts away from perhaps the most basic general consequence of this motivation. Namely, an agent with ego utility prefers to receive her information in the most flexible and manipulable manner possible, so that she can tailor information acquisition finely to achieve a favorable self-image. In particular, I prove that—holding total informativeness constant—the agent prefers to receive information in smaller pieces.

The model in this paper is intended to be a relatively general model of distortions in behavior due to ego utility, with many potential economic applications. As an illustration of some of the key results, consider two sides of an important decision, investing in the stock market. A small investor has to decide whether to invest in stocks and, if so, how much to trade. Investing in stocks is more ambitious than not investing, and trading much is more ambitious than trading little. In addition, picking stocks is more informative about investing skill than not doing so, but if the agent follows the market, to a first approximation she learns the same about her skill whether or not she trades a lot. If investors attach ego utility to thinking that they are good enough to enter the stock market, and to trade a lot, overconfidence develops: When asked, too many investors would claim that they would be successful if they chose the ambitious goal. Behavior is more complicated. Because entry is more informative than investing in safer assets, because of self-image protection some decision makers with positive beliefs do not enter the stock market. But because trading a lot is not much more informative than trading little, those who enter act on their overconfidence and actually trade too much. Thus, ego utility can lead to underparticipation and overly active trading conditional on participation.

Other applications include small businesses, project choice by managers, career or educational choice, job search, and the extrinsic and intrinsic motivation of employees. Two of these examples are discussed in Section 4.

The paper is organized as follows. Section 2 presents the setup and comments on the model. Section 3 analyzes the model and discusses the robustness of the main results. Applications are taken up in Section 4. Section 5 gives detailed evidence for both the assumptions and the results of the model, and discusses related literature. Section 6 concludes. All proofs are in Appendix B.

2. A Model of Behavior with Ego Utility

The model starts from a simple premise: People have *ego* about what activity they see themselves capable of succeeding in. In developing the general model, I will largely use an application-neutral language, but a specific setting is useful to keep in mind for motivation. Consider the decision between having a salaried job (e.g., working at a record shop) and starting one's own business (owning a record store). On the one hand, it probably takes more business skill to manage one's own music dealership than being an employee in someone else's. On the other hand, the potential returns are also higher, so this career choice makes sense if one is good enough. In addition to caring about financial outcomes, a person might derive pleasure from thinking that she belongs to the distinguished group of people who could have a successful business.

Formally, the model has two conceptually different kinds of activities, separated into different types of decision periods: Period 0, an initial learning period in which the agent can gather information about herself, and periods 1 and 2, financial choice periods in which decisions directly affect instrumental outcomes. The basic timing of the decision problem is illustrated in Figure 1. The agent enters the decision problem with ability q. The distribution of q in the population, and the agent's prior, is $N(\mu_a, \sigma^2)$. In period 0, the agent can sequentially draw signals about q. In period 1, she makes a financial decision a_1, and—depending on

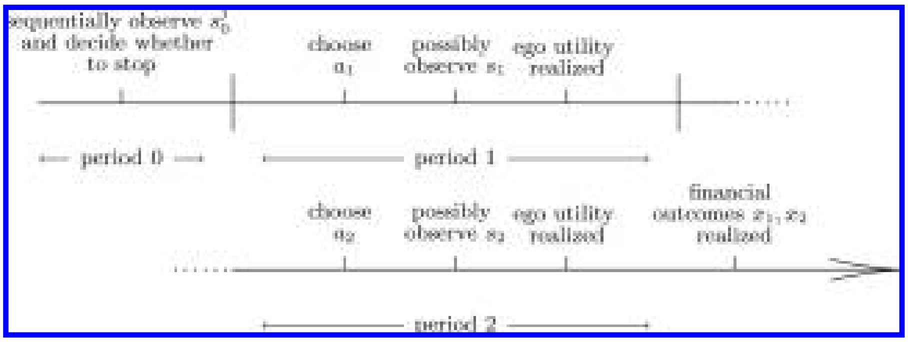

FIGURE 1. Timing.

the variant of the model—may observe a random variable s_1 whose distribution depends on q and which determines her financial outcome. Next, she derives ego utility from her momentary beliefs. In period 2, the same three steps repeat themselves. Finally, the two financial outcomes $x_1(s_1, a_1)$, $x_2(s_2, a_2)$ are realized.

I begin by fleshing out the details of the above decision problem. In period 0, the initial learning period, the agent can observe an arbitrary, possibly infinite, number of free signals $s_0^j = q + \varepsilon_0^j$, where $\varepsilon_0^j \sim N(0, \sigma_s^2)$, and the ε_0^j are independent. After each s_0^j, the decision maker can decide whether to keep drawing signals or to stop and move to period 1. The signals are available only in period 0, not in periods 1 and 2. In Section 3.5, I discuss how the results would be different if the learning opportunity reappeared later, and if there was no early learning opportunity. The specific assumption that the agent can draw an infinite number of signals does make some predictions more extreme, but it is not important for any of the qualitative results in the paper.

The assumption about learning in period 0 captures the idea that people enter most decision situations having a notion about how well they will do in them, but how much and what they know depends at least partially on themselves. At the most literal level, the assumption corresponds to a voluntary search for external signals of ability. But the model applies equally easily to an internal thought process in which people recall, consider, and interpret relevant facts about themselves and the world and can voluntarily stop their deliberation.[2] A person might, for example, know that she has good social skills, but it could take a lot of thought to decide whether this is important for being a good employee, which is what she may really care about.

2. This view finds support in some psychology research on memory. Greenwald (1980), for example, takes the view that the ego influences the memory process, comparing it to a totalitarian political regime and citing Orwell's 1984: "Who controls the past," ran the Party slogan, "controls the future: Who controls the present controls the past." For evidence that people use their memory to improve their self-image, see Murray and Holmes (1994).

In each period $t \in \{1, 2\}$, the agent chooses an action a_t from two options: $a_t \in \{0, 1\}$. The "unambitious" option 0 has a certain payoff of 0 in each period, while the "ambitious" option 1 pays off either -1 or 1. The risky option is ambitious because the probability of success in it is increasing in the agent's ability q. Specifically, a signal $s_t = q + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma_s^2)$ is generated independently in each period t, and the ambitious undertaking succeeds in that period ($x_t = 1$) if $s_t > 0$. For simplicity, I use the convention that if the agent is indifferent, she chooses the unambitious option.

The role of the financial decision periods is twofold. First, they introduce a cost for manipulating one's ego, because inaccurate self-views lead to lower financial payoffs. Second, the choices and beliefs in these periods can be used to study how the desire to manage one's ego interacts with economic decisions, one of the main interests in this paper.

I now turn to a description of the agent's preferences in the above setting. She cares about the standard financial outcomes x_t, and for simplicity I assume that she is risk-neutral with respect to these outcomes. Crucially, however, her utility function also includes a non-standard component, ego utility. Ego utility in period t depends on the decisionmaker's beliefs about q at that time.[3] Let F_t denote the cumulative distribution function of the agent's beliefs about q at the end of period $t \in \{0, 1, 2\}$. For methodological and other reasons detailed in Section 5, I assume that F_t is deduced from the agent's prior and available information in a Bayesian way. Following Caplin and Leahy's (2001) formulation of beliefs-based utility, I assume that the agent is an expected-utility maximizer over the enriched payoff space that includes financial outcomes and ego. Specifically, her von Neumann-Morgenstern utility function takes the form

$$wu(F_1) + x_1(s_1, a_1) + wu(F_2) + x_2(s_2, a_2). \tag{1}$$

w parameterizes the importance of ego utility. There is no discounting.

Like most economic models, this theory requires some structure on the utility function—in this case, the ego utility part. The key assumption is that the decisionmaker likes to believe that she is good enough to try the ambitious option. More precisely, the agent's ego utility is 1 if she believes it makes financial sense to go for option 1, and it is zero otherwise. Given risk neutrality in financial payoffs, the ambitious option is financially superior if the subjective probability of getting the outcome 1 is at least one-half. Because for beliefs F the subjective distribution of s_t is centered around the mean μ_F of F, ego utility is 1 if and only if $\mu_F > 0$.

3. In Baumeister's (1998) language, ego utility would correspond to the "reflexive self," the part of our psyche that examines ourselves. Instrumental utility motivates the "executive self," which is responsible for making decisions. In the previous decision problem, both of these selves are evoked at the same time.

This step-function form for ego utility is rather special. Section 3.6 shows that some of the main results in the paper depend not on the step-function nature of u, but on its tight link to the decision problem at hand. That is, in the above formulation, ego utility is 1 if, according to current beliefs, $a_t = 1$ is the financially appropriate choice, and ego utility is 0 if $a_t = 0$ is the financially appropriate choice. As long as this kind of property is preserved in more complicated settings, the overconfidence result and the time patterns of self-image protection and enhancement generalize to those environments. Results on how overconfidence, self-image protection, and self-image enhancement vary with the level of the agent's beliefs also rely on the convexity of ego utility over low beliefs, and the concavity of ego utility over high beliefs.[4] A step-function ego utility captures the qualitative features of such preferences. Section 5 discusses some evidence consistent with this assumption.

Assumption 1 collects the assumptions on the agent's decision problem and ego utility.

ASSUMPTION 1. $a_t \in \{0, 1\}$, $s_0^j = q + \varepsilon_0^j$, $s_t = q + \varepsilon_t$, $\varepsilon_0^j, \varepsilon_t \sim N(0, \sigma_s^2)$, and

$$x_t(a_t, s_t) = \begin{cases} 0 & \text{if } a_t = 0 \\ 1 & \text{if } a_t = 1 \text{ and } s_t > 0 \\ -1 & \text{if } a_t = 1 \text{ and } s_t \leq 0. \end{cases}$$

Ego utility takes the form

$$u(F_t) = \begin{cases} 1 & \text{if } \mu_{F_t} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

In addition, I will consider various assumptions on the choices that lead to observing s_t:

ASSUMPTION A. s_t is never observed.

ASSUMPTION B. s_t is always observed.

ASSUMPTION C. s_t is observed if and only if $a_t = 1$.

Assumption A applies best to situations (e.g., retirement investment) where performance outcomes are observed so long after the decision that they have little impact on current ego-utility concerns. Assumption B is most relevant when feedback about the quality of chosen as well as unchosen actions is fast and

4. By the law of iterated expectations, the decision maker's mean beliefs do not change in expectation when she receives new information. Therefore, by Jensen's inequality, the part of the ego utility function that is concave in the decisionmaker's mean beliefs is associated with a preference to avoid information, and the convex part is associated with a preference to acquire information.

accurate (as when the amount of investment in a short-term scalable project has to be decided). Assumption C, perhaps the most broadly relevant one of the three, is appropriate when the agent's ability is only tested in the ambitious task (as in the small business example).

3. Results

The agent's strategy in period 0 can be fully described by functions $a_0^j : \mathbb{R} \to \{0, 1\}$, $j = 0, 1, 2, \ldots,$ where $a_0^j(\mu) = 1$ if and only if the agent chooses to continue sampling when she has drawn j signals and has mean beliefs μ.[5] I start with establishing existence of optimal strategies.

PROPOSITION 1. *Suppose Assumption 1 and one of Assumptions A, B, or C hold. Then, an optimal (continuation) strategy exists for any initial or interim beliefs in period 0.*

Once the agent has stopped acquiring information in period 0, and has beliefs F, her optimal strategy in periods 1 and 2 can be solved by backward induction. Let her expected utility under this optimal strategy be $U(F)$. For example, under Assumptions 1 and A, she acquires no information in periods 1 and 2, so $U(F)$ takes the following simple form:

$$U(F) = 2wu(F) + 2\left(\max_{a\in\{0,1\}} E_F[x_1(a, s_1)]\right). \tag{3}$$

Denote by $V(F_0^j)$ the agent's maximum expected utility given beliefs F_0^j after having observed j signals in period 0. Let $F_0^{j+1}(F_0^j, s_0^{j+1})$ be the beliefs induced from F_0^j by the signal s_0^{j+1}. The agent's information-acquisition decision in period 0 is a dynamic optimization problem with a Bellman equation

$$V\big(F_0^j\big) = \max\left\{U(F_0^j), E_{F_0^j}[V(F_0^{j+1}(F_0^j, s_0^{j+1}))]\right\}.$$

Completely describing the solution to this problem is extremely difficult, and not a goal of this paper. The sections below establish the properties needed for the results.

3.1. Overconfidence in Beliefs

I begin by showing that through learning about herself in period 0, the agent acquires overconfident beliefs regarding the financial choice appropriate for her.

5. Given the normality assumptions on the prior and the signals, j and μ fully describe the agent's beliefs (because the variance of her beliefs only depends on the number of signals observed).

This result is established in two parts. First, the following proposition shows that if the agent's mean beliefs about q are nonpositive, she continues to draw signals.

PROPOSITION 2. *Suppose Assumption 1, and one of Assumptions A, B, or C, holds, and the agent holds beliefs F_0^j in period 0. If $\mu_{F_0^j} \leq 0$, she chooses to sample the next signal, s_0^{j+1}.*

An agent who has a low self-image is eager to find out more about herself for both self-image and instrumental reasons, and thus faces no tradeoffs. On the self-image side, information raises her beliefs above zero with positive probability, increasing her ego utility. And on the instrumental side, information can be useful for making better financial decisions. Formally, returning to Assumptions 1 and A, for any F_0^j with $\mu_{F_0^j} \leq 0$ we have $E_{F_0^j}[u(F_0^{j+1}(F_0^j, s_0^{j+1}))] > u(F_0^j)$, because mean beliefs move above zero with positive probability as a result of observing s_0^{j+1}. In addition,

$$\max_{a(\cdot)} E_{F_0^j}\left[x(a(s_0^{j+1}), s_1)\right] \geq \max_a E_{F_0^j}[x_1(a, s_1)],$$

because the agent can ignore s_0^{j+1} in choosing her actions. Combining these two inequalities and using equation 3 yields

$$U(F_0^j) < E_{F_0^j}\left[U(F_0^{j+1}(F_0^j, s_0^{j+1}))\right] \leq E_{F_0^j}\left[V(F_0^{j+1}(F_0^j, s_0^{j+1}))\right],$$

so the decision maker chooses to sample further.

Although this intuition for Proposition 2 may seem straightforward, there is a subtlety in the proof related to the "option value" discussion in Section 3.7. Namely, if for some reason the agent prefers the signals available in period 0 to those available in later periods, to wait for those signals she may be willing to live with negative beliefs in period 0. Because signals in periods 0, 1, and 2 are perfect substitutes, this consideration is absent in the current model.

As the second ingredient for the overconfidence result, Proposition 3 shows that for positive mean beliefs in period 0, the analogue of Proposition 2 does not hold. Specifically, in this range, the agent stops acquiring information with positive probability.

PROPOSITION 3. *Suppose Assumption 1 and one of Assumptions A, B, or C hold. For any beliefs F_0^j in period 0, in the agent's optimal continuation strategy, she perfectly learns q with probability strictly less than one.*

To understand the intuition for Proposition 3, suppose once again that Assumption A holds. Because the agent does not learn more about herself after period 0, in periods 1 and 2 she chooses the financially superior option given her final beliefs in period 0. Therefore, sampling more information only affects her utility if she ends up with mean beliefs below zero (otherwise, both her ego utility and her financial choices remain the same). Being forced to conclude that $q \leq 0$ decreases the agent's ego utility from one to zero, but increases her instrumental utility because she makes a better financial decision. Proposition 3 proves that with positive probability, the former consideration outweighs the latter.

Given Propositions 2 and 3, the beliefs at the end of period 0 have the property that all agents with below-zero mean beliefs know their types with certainty, though this is not the case for those with above-zero mean beliefs. As a consequence, some of those holding positive beliefs in fact have $q \leq 0$, and no agent with negative beliefs has $q > 0$. Therefore, more agents hold positive mean beliefs than there are positive types. While I turn to actual behavior below, this means that relative to the actual distribution of abilities, too many people rationally and honestly believe that from a financial point of view they should choose the ambitious option. In other words, even though agents are Bayesian and thus unbiased in their beliefs about q, their average beliefs regarding the appropriate financial action for them are biased upwards.[6]

Propositions 2 and 3 not only imply overconfidence, but also a systematic relationship between the level of the decisionmaker's beliefs and the bias in them. In particular, confident agents (those holding positive beliefs) are overconfident on average, whereas less confident agents are unbiased.

It might seem surprising that agents who process information in a Bayesian way can be overconfident in any measure. In fact, the decision maker may appear non-Bayesian in that in forming her beliefs, she does not condition on the fact that she tends to stop information collection when mean beliefs are positive. But whatever signals she has observed that led her to stop information acquisition, the fact that she stopped provides no information about the state of the world, because she stops acquiring information after this set of signals in all states.[7] The counterintuitive conclusion is that before period 0, the decision maker knows she will too often end up with favorable beliefs, but it is still rational for her to hold those beliefs whenever she does.

6. The results so far imply only that the decision maker will be overconfident in the defined sense at the end of period 0, but the overconfidence clearly persists through periods 1 and 2 in all variations of the model.

7. Formally, let the stopping rule be the random variable $h(S)$, which takes the value of 1 if the agent stops acquiring information after having observed the set of signals S, and zero otherwise. Suppose S is a set of signals after which the agent stops sampling signals. Then, by Bayes' rule, the density f satisfies

$$f(q|S, h(S) = 1) = \frac{f(q) f(S|q) \Pr(h(S) = 1|q, S)}{f(S) \Pr(h(S) = 1|S)} = \frac{f(q) f(S|q)}{f(S)} = f(q|S),$$

because $\Pr(h(S) = 1|q, S) = \Pr(h(S) = 1|S) = 1$. So the stopping rule does not affect updated beliefs about ability.

Knowing the specific decision problem, it is of course easy to design measures for which overconfidence would not be observed. For example, too few agents might believe that their expected payoff in the ambitious option is at least 1/2. In practice, however, most measures of people's beliefs that studies elicit and that one might be interested in invoke strong ego utility in that people want to believe they stand higher on these measures. Thus, a positive bias is likely to be observed in most cases.

The only measures in which a Bayesian model is in general unbiased are those that are linear in beliefs about the underlying state of the world. It is worth emphasizing, however, that even this property of Bayesian updating depends strongly on the assumption of correct priors. If people have, say, dispersed but on average still correct priors, overconfidence results even for decision makers' probabilistic beliefs: Agents with too-low priors seek out more information than agents with too-high priors, increasing average beliefs relative to the priors.

3.2. Overconfidence in Actions

I now proceed to the analysis of the agent's financial decisions, and how they relate to her beliefs about q. Assume first that Assumption A or B is satisfied—the agent's financial decision cannot change the information she receives, and hence the ego utility she experiences, in periods 1 and 2. She then chooses the option she believes has the higher financial return. Combined with the result established previously that her beliefs are overconfident, this implies overconfidence in her actions.

If Assumption B (that s_t is observed with either choice) holds and the agent chooses $a_1 = 1$, there is a positive probability that she gets a signal low enough to move her mean beliefs about q below zero. That is, she learns she is probably not good enough to be ambitious "the hard way" instead of from free information at the beginning. Because a decision maker with negative beliefs knows her ability with certainty and hence does not change her beliefs in period 1, the probability that she learns $a_1 = 1$ is optimal for her is zero. Because some agents update their beliefs regarding the optimal action downward, whereas no agent updates her beliefs upward, beliefs and choices become less optimistic on average over time. Intuitively, because the agent starts period 1 with overconfident beliefs, the unavoidable information she receives tends to make her more realistic.

3.3. Self-Image Protection and Rational Hypocrisy

I turn to the agent's behavior under Assumption C, that s_t is observed if and only if $a_t = 1$. This means that in order to see whether she would succeed in the ambitious activity, the agent has to try it. One of the new effects that surfaces in this variant of the model is "self-image protection" in period 1.

PROPOSITION 4. *Suppose Assumptions 1 and C hold. Then, there exist constants* $\mu_1^j \geq 0$, $j = 0, 1, 2, \ldots$ *such that if the agent enters period 1 with beliefs* F_0^j *having drawn* j *signals in period 0, she chooses* $a_1 = 1$ *if and only if* $\mu_{F_0^j} > \mu_1^j$.

When the agent has good, but uncertain, views of herself, being ambitious carries the risk of finding bad news about her ability and thereby decreasing her ego utility. As a result, she may not take on the ambitious task, even though it is superior by financial standards.[8]

This variant of the model shares the property of previous versions (Assumptions A and B) that more people think they are good enough for the ambitious option than is warranted by the population distribution of abilities. The curious situation here, however, is that these beliefs are not necessarily translated into action, creating a kind of hypocrisy. We are all familiar with this kind of individual—he is happy to advertise his ability in a domain (and even to criticize others for their performance), but when challenged to show how it is done, his arrogance suddenly fades. In fact, there is a negative relationship between beliefs and actions in the following sense.

PROPOSITION 5. *Suppose Assumption 1 holds.*

1. Under Assumption A, B, or C, the agent's expected ego utility $(E[u(F_1) + u(F_2)])$*, and thus average overconfidence, increases with* w.

2. If Assumption C holds, then as $w \to \infty$*, the probability of participation in the ambitious task approaches zero.*

Because ego utility is responsible for both the overconfidence in beliefs and the underconfidence in financial choices, as it becomes more important, both effects tend to get stronger.

Proposition 5 yields a (possibly testable) out-of-sample prediction of the model depending on whether s_t is observed symmetrically after each action. Consider the effect of an increase in monetary stakes on the agent's behavior, which is equivalent to a decrease in w. When s_t is symmetrically observed (Assumption A or B), participation in the ambitious task decreases. Intuitively, as stakes increase, the agent prefers to hold a more accurate self-view, decreasing her confidence in beliefs and behavior. When s_t is only observed after $a_t = 1$ (Assumption C), participation in the ambitious task tends to increase (although this is now not generally true for the entire range of stakes). Intuitively, one way in which the agent manipulates her ego is to avoid the ambitious option even when it is financially appropriate. As stakes increase, she is less willing to take this financial sacrifice.

8. The converse of the same idea and proof will be used below to establish the self-image enhancement motive in period 2, that some agents with uncertain negative mean beliefs choose the ambitious activity. The only reason this does not arise in period 1 is that the agent never starts the activity with uncertain negative beliefs.

In a corresponding model without ego utility ($u(\cdot) \equiv 0$), participation in the ambitious task does not depend on the stakes.

In a sense, the self-image–protecting distortion in financial behavior is analogous to the ego-protecting rejection of free signals in period 0, an effect also present in previous variants of the model. But there is an interesting difference between the two ways of self-image protection. Being unambitious with positive beliefs is a sign of "insecurity" in those beliefs: If the agent has very high beliefs, she prefers being ambitious because it is likely to yield both a financial outcome of 1 and ego utility of 1. An agent who refuses to consider free information, however, is not necessarily insecure—in fact, it might be the highest types who are confident enough to ignore free signals.[9]

3.4. Self-Image Enhancement

To complete the analysis of the basic model, I consider the agent's behavior in period 2 under Assumption C.

PROPOSITION 6. *Suppose Assumptions 1 and C hold. Then, there exist constants* $\mu_2^j > 0$, $j = 0, 1, 2, \ldots$ *such that if the agent enters period* 2 *with beliefs* F_1 *having drawn* j *signals in periods* 0 *and* 1, *she chooses* $a_1 = 1$ *if and only if* $\mu_{F_1} > \mu_2^j$ *or* $-\mu_2^j < \mu_{F_1} \leq 0$.

When the agent has uncertain negative mean beliefs in period 2, she faces a tradeoff between ego and instrumental utility. The ambitious option has a lower expected financial payoff, but it provides a chance for ego utility to increase to 1, which is impossible with the unambitious option. For mean beliefs close to zero, the financial consideration is relatively weak, so the decision maker engages in "self-image enhancement": She chooses the ambitious activity despite expected financial losses, hoping for a good financial outcome to improve her self-image.

Because the decision maker can have uncertain positive beliefs in either of the two periods but can only have uncertain negative beliefs in period 2, self-image protection can arise in either period, but self-image enhancement can only happen in the second. Although this prediction is extreme and relies on the assumption that the agent can view an arbitrary number of signals in period 0, the basic intuition is likely to hold in many settings. When a person has an abundance of opportunities to learn about herself before engaging in an activity, she is likely to stop learning when she no longer wants information. Thus, she enters an average new situation

9. In contrast to the current paper, psychologically it seems possible for the very hypocrisy identified above to make a person uncomfortable about herself. If someone likes to think of herself as a prudent financial decision maker—as opposed to just someone who is good enough to be ambitious—she might feel bad about doing something financially suboptimal. This might mitigate hypocrisy, but it is unlikely to be strong enough to eliminate it altogether.

with a self-image protection motive. In the course of doing a skill-sensitive new activity, she may lose her positive self-view and as a result engage in self-image enhancement to reestablish it.

3.5. Modifications of the Basic Model

To gain further insight into the forces driving the results of the previous sections, I briefly discuss how certain changes to the basic model modify its predictions. First, in contrast to the model's assumption that free signals are unavailable after period 0, there are dimensions of ability in which cheap learning opportunities are available at any time.[10] This can be modeled by assuming that the agent can sample an arbitrary number of free signals in all periods, not only in period 0. Then, she never wants to distort her financial choices for self-image-enhancement reasons—she would rather sample more free signals. However, under Assumption C, she might still distort her choice of tasks to protect her self-image. For this reason as well as the noted tendency to enter many situations with a self-image–protection motive, the theory predicts more self-image protection than self-image enhancement in instrumentally relevant choices in the real world.

Second, at the other extreme, there are environments where cheap learning opportunities are rarely available, or (as in the example at the beginning of Section 3.7) the agent does not want to take advantage of them. She is then unable to manipulate her beliefs before the actual choice problem, so it is not in general true that she enters period 1 with a self-image protection motive and only later engages in possible self-image enhancement. In addition, if she can manipulate her instrumental choices to improve her self-image, in contrast to the basic model overconfidence develops rather than diminishes during the activity.

Third, although such situations seem to be rare, the ambitious option is sometimes less informative about the relevant ego parameter than the unambitious one. Then, when the agent thinks instrumental considerations warrant the ambitious option, self-image protection gives her an extra reason to choose it. Conversely, when she has bad beliefs, she wants to be unambitious for both instrumental and ego reasons. Therefore, beliefs and actions coincide. And because there is no chance for overconfident beliefs to be proven wrong once established, such beliefs are very persistent in this setting.

An example consistent with the notion that the ambitious option is less informative comes from Brandenburger and Nalebuff (1997). They note that some people believe they are indispensable for their firm, and thus never take a vacation—never finding out whether they are really so important.

10. For many applications, the situation is between the two extremes. A person might always have some free information she can draw on to improve her self-image, but once participation in a skill-sensitive activity has started, acquiring information through that channel might be the easiest and fastest way to enhance one's self-image.

3.6. Robustness

This paper started with the premise that people derive pleasure from positive beliefs about their abilities and traits. Much of the psychological evidence in Section 5 centers around this general premise. To incorporate ego utility into an actual decision problem, however, I have made more specific assumptions regarding the ego utility function u. Section 5.1 summarizes some evidence that u captures key features of people's self-image management motives. This section, in turn, discusses the theoretical robustness of the results in the paper. It shows that the two-point range of u is not what gives the model a lot of its mileage.

Consider the following assumptions.

ASSUMPTION 2. Suppose $a_t \in \mathbb{R}$, $x_t \in \mathbb{R}$, $s_0^j = q + \varepsilon_0^j$, $s_t = q + \varepsilon_t$, $\varepsilon_0^j, \varepsilon_t \sim N(0, \sigma_s^2)$, and $x_1(\cdot) = x_2(\cdot)$ is an arbitrary function. After taking an action $a_t \in \mathbb{R}$, the agent observes $q + \eta_t$, where $\eta_t \sim N(0, \sigma_{a_t}^2)$. For any $a \in \mathbb{R}$, σ_s^2/σ_a^2 is an integer. Assume that for any strategy of the agent and any period t, the probability that for a belief F_t she reaches $a^*(F_t) \equiv \text{argmax}_{a\in\mathbb{R}} E_{F_t}[x_t(s_t, a)]$ is nonempty- and single-valued is 1, and $u(F_t) = a^*(F_t)$ wherever this is the case.

The actions taken in periods 1 and 2 are now on the real line instead of binary, and the financial outcomes depend in an arbitrary way on s_t and a_t. If the agent takes action a_t, she observes a signal of q, and the variance of this signal depends on a_t.[11] The role of the extra assumption that σ_s^2/σ_a^2 is an integer will be made clear in Section 3.7. Finally, ego utility is identical to what action the agent thinks she should take from a financial point of view.[12] That is, the agent likes to think that she is capable of taking higher actions (or, equivalently, actions are ordered from low to high in terms of generating ego utility). For instance, there may be many potential language courses available to a student, some of which are too easy for her, some of which are too hard, and one of which is best in terms of improving her language skills. The more advanced she believes is the appropriate course for her, the better the student feels about her ability.

With these ingredients, many of the main points of the paper generalize. For the statement of Proposition 7, denote by δ_q the probability measure that assigns unit mass to q.

11. The basic model herein is equivalent to the case when the variance of this noise term is either ∞ or σ_s^2. Because financial outcomes are realized only at the very end, it makes no difference to the analysis that the agent observes a different signal than that determining financial performance.

12. For simplicity, the optimal action is assumed to be unique. Analogous definitions and results would hold, but would be more difficult to state, if this was not the case.

PROPOSITION 7. *Under Assumption 2, the following hold.*

1. *The average beliefs of the agent about the optimal action are overconfident:*
$$\frac{E[a^*(F_1) + a^*(F_2)]}{2} \geq E[a^*(\delta_q)].$$
2. *If all actions are uninformative ($\sigma_r = \infty$ for all r), her actions are also overconfident: $E[a_1 + a_2]/2 \geq E[a^*(\delta_q)]$.*
3. *If higher actions are more informative (σ_r is decreasing in r), in the first period the agent almost always chooses a lower action than she believes is financially optimal: $a_1 \leq a^*(F_0)$ with probability 1. This is not necessarily the case in the second period.*

Given the generality of the decision problem in Assumption 2, some key results in the paper hold for virtually any ego-utility function u. Part 1 says that the agent develops overconfident beliefs in the sense that her average beliefs about the appropriate action are higher than that warranted by the actual distribution of ability—that is, than the average beliefs she would have if she knew her ability with certainty. Part 2 says that this overconfidence can translate into overconfidence in actions. And Part 3 says that if higher actions are more informative, the agent protects her self-image by taking lower actions in period 1 than would be financially optimal, but the opposite may be the case in period 2.

Rather than the form of u, the crucial assumption in the model is that $u(F) = a^*(F)$, which establishes a tight link between ego utility and the instrumental side of the problem being studied. Although much stronger than just assuming the existence of ego utility, positing a relationship between ego and financial prospects seems reasonable: The agent not only likes to believe that she is "good," but also that she is "good enough to be ambitious."[13]

The inequalities in Proposition 7 are weak inequalities, and the results are more interesting when—as in the basic model—they hold strictly. The proposition is not true in general with strict inequalities; for example, if ego utility was linear in beliefs, it would be optimal for the agent to acquire all information, and not to distort any of her financial decisions. Loosely speaking, the results should be strict if the ego-utility function has both strict convexities and strict concavities in the relevant range of possible beliefs, but the location of these nonlinearities is not important.

13. Of course, a person's ego utility can depend on much more than whether she thinks she is good enough to be ambitious. For example, she might feel better if she is more certain that she would make money with the ambitious option. That is, ego utility might depend on the variance of beliefs as well. The thrust of the model would be very similar if these more complicated considerations were included.

Although Proposition 7 generalizes key results in the paper, it does not cover all points made in the basic model. Results on how the agent's overconfidence and behavior depend on the level of her beliefs also rely, broadly, on the ego utility function being concave for higher beliefs and convex for lower ones. Without this assumption, it would not be true that more confident agents are more biased than less confident ones. Also, it would not be true that self-image protection is associated with higher beliefs, and self-image enhancement with lower ones. Finally, it would not be the case that with dispersed but on average correct priors, overconfidence still develops.

One potential problem with assuming a link between ego utility and the financial options is that it only makes sense within a given decision problem. A person's ability may affect many decisions in her life, and do so in many different ways. Thus, if she has a single ego-utility function over her beliefs about ability, this may not be tightly linked to any one instrumental decision problem; however, people often consider decisions one-by-one, disintegrated from other decisions. Instead of focusing on our beliefs about ability and all possible decisionmaking problems at all times, we develop precise beliefs and strategies only when confronted with a specific choice. And in doing so, we may be motivated to develop a positive self-image within the context of the decision problem.

3.7. Option Value for Ego Utility

This section shows that a decisionmaker with ego utility prefers to acquire information in small pieces, a consideration ignored in the analysis so far. As a motivating example, consider a modification of the basic model with Assumptions 1 and C. Suppose that in period 0, the decision maker has only two choices after receiving her first signal: She can either immediately learn q, or receive no more information in period 0. Somewhat, surprisingly, it is then no longer in general true that with negative mean beliefs the decision maker always samples information. In particular, if w is very large and her mean beliefs are close to zero, she prefers not to learn q. To maximize ego utility—which is essentially all she cares about—she instead first engages in the ambitious activity, and does the same in period 2 if and only if her mean beliefs are still below zero. This way, the probability that she ends up with ego utility equal to 1 is approximately one-half in period 1, and strictly greater than one-half in period 2—whereas it would be approximately one-half in both periods if she found out q.

What drives the agent's reluctance to find out her type is her fear of finalizing her beliefs. Even though learning her type can only improve her momentary self-image, it carries an important risk: If she finds that $q \leq 0$, her ego utility will with certainty be low forever. To manipulate her beliefs more easily, she prefers to collect information in smaller pieces, and waiting for period 1 allows her to do that. Leaving beliefs uncertain therefore has an *option value* in managing self-image.[14]

The basic model abstracts away from the above consideration by assuming that the signals in periods 0, 1, and 2 have the same variance and thus are perfect substitutes. And option-value considerations do not affect the general results in Section 3.6 because of the assumption that the variance of signals in period 0 is a multiple of the variance of later signals—so that period-0 signals are substitutes for later signals.

To formalize the option-value motive, I modify the setup of Section 2 to allow for a comparison between different information structures in period 0. Instead of assuming that the decision maker can choose whether to observe s_0^j one-by-one, suppose that the same signals are available in groups. More precisely, she first chooses whether to observe k_1 signals, where k_1 is a positive integer or infinity. If k_1 is finite and the agent chooses to observe these signals, she decides whether to observe the next k_2 signals, and so on. An information structure can then be described by a sequence $(k_1, k_2, \ldots, k_i, \ldots)$, where the sequence is either infinite, or its last element is equal to infinity. This leads to the following proposition establishing that the decision maker always wants smaller pieces of information.

ASSUMPTION 3. Suppose $a_t \in \mathbb{R}$, $x_t \in \mathbb{R}$, $s_0^j = q + \varepsilon_0^j$, $s_t = q + \varepsilon_t$, $\varepsilon_0^j, \varepsilon_t \sim N(0, \sigma_s^2)$, and $x_t(s_t, a_t)$ and $u(\cdot)$ are arbitrary functions. After taking an action $a_t \in \mathbb{R}$, the agent observes $s_t + \eta_t$, where $\eta_t \sim N(0, \sigma_{a_t}^2)$.

PROPOSITION 8. *Suppose Assumption 3 holds, and $k_i = k_i' + k_i''$. Then, the agent prefers the information structure $(k_1, \ldots, k_{i-1}, k_i', k_i'', k_{i+1}, \ldots)$ over the information structure $(k_1, \ldots, k_{i-1}, k_i, k_{i+1}, \ldots)$.*

4. Two Applications

This section discusses two possible economic applications of the model. I first elaborate on the small-business example used to motivate the model, and then discuss a more complex set of applications. There are of course standard explanations for most of the phenomena discussed herein. An advantage of the current paper is that it provides a single framework consistent with many observations.

4.1. Small Businesses

As mentioned previously, the paper has a straightforward application to small businesses. Prospective entrepreneurs might derive utility from the thought that

14. Note the interesting distinction between this and standard option value: Whereas the standard option-value consideration would induce an agent to take more informative actions, option value for ego utility often leads her to prefer less informative ones.

they have the skills necessary to create a lucrative business rather than just perform satisfactorily in a salaried job. If this is the case, the model predicts that people will be on average overconfident about their ability to run a business if they had one. At the same time, managing an enterprise is often more informative about business skill than being employed in an average job. Therefore, due to the self-image protection motive, some of those who believe they should (from a financial point of view) own a business in fact do not take on the challenge. A person who tries her luck and does worse than expected may respond in two ways. If her performance is so bad that she loses her positive self-image, to regain confidence she may stay in the unsuccessful enterprise for too long. But if her performance is not as bad, she may choose the unambitious option in the second period despite still having positive mean beliefs: To escape with her ego intact, she prematurely closes her business.

4.2. Motivation of Employees

The organizational setting is perhaps the most fruitful application of ego utility, with both straightforward utilization of the model in this paper and more complex further questions.

The ego-utility model reconciles a pair of seemingly contradictory facts from the organizational-economics literature. As discussed in Section 5, employees are in general overconfident about their job-related skills, most rating themselves well above comparable coworkers. In a standard model, confident workers should be enthusiastic about performance bonuses, as these bonuses benefit the best workers. Yet employees seem to dislike performance-based incentives (Milkovich and Newman 1987). This apparent contradiction is consistent with the self-image-protecting divergence between beliefs and actions identified in Section 3.3: Employees have convinced themselves that they are good, and they do not want to find out the opposite from an informative pay structure.

How to design incentives to take advantage of the ego factor and minimize its distortions is an interesting area for future research. Indeed, in several ways ego utility could be a source of "intrinsic motivation," people's willingness to work hard without explicit economic incentives to do so. For instance, if many employees in many situations have a self-image protection motive (as predicted by this paper), the firm wants to create conditions under which working hard jams signals about ability. And more directly, in order for a person to feel that she has made a good choice regarding her work, she may want to believe that the job she is performing is intrinsically worthwhile.

Even more interestingly, substantial experimental evidence indicates that extrinsic incentives crowd out other kinds of motivation, and therefore they might not be as desirable as standard principal-agent models predict.[15] Psychologists explain the evidence by a "rationalization" effect, that only people not being paid to perform a task need to justify their involvement in it by convincing themselves that it is intrinsically worthwhile. An extension of the ego utility model can naturally accommodate this kind of crowd-out effect. Suppose two kinds of information affect the agent's ego utility: Whether it is important for one's job to be enjoyable, and whether her job is enjoyable.[16] She prefers her job to be enjoyable if that is important, but prefers enjoyment to be unimportant if her work is not fun. With no financial incentives, the agent wants to convince herself that it is important to enjoy one's job and that her job is enjoyable, creating an intrinsic motivation to work hard. But if she gets financial incentives to exert high effort, she will want to do so even if the job is not so enjoyable. Expecting this, she might change her information-collection strategy to convince herself that enjoying one's job is not so important. Then, she does not have an incentive to convince herself that the job is enjoyable, undermining intrinsic motivation. In other words, expecting that she would want to do the job for the money no matter what leads the agent to believe that it is not important to have other reasons to do it, which in turn leads her to believe that there is no other reason to do it.

Of course, it is unlikely that extrinsic motivation always undermines other kinds of incentives. A theoretical model of the this kind might delineate the circumstances in which we can expect this to be the case, a question that is not satisfactorily addressed in either the theoretical or empirical research on this topic.

5. Psychological Evidence and Related Literature

This section reviews psychological evidence for the ego-utility model, and discusses related economic theories. I consider two categories of evidence: Those relating to the assumptions of the model and those relating to its predictions.

5.1. Evidence on the Assumptions of the Model

The most direct evidence that people have preferences over their beliefs about themselves comes from the literature on cognitive dissonance (Festinger 1957). In the main experimental paradigm for studying cognitive dissonance, subjects are induced to perform an act they consider morally or ethically wrong. This is

15. For basic crowdout experiments, see for example Deci (1972) and Lepper, Greene, and Nisbett (1973). For discussions of these experiments and the desirability of extrinsic incentives, see Baker, Jensen, and Murphy (1988), Deci (1972), and Bénabou and Tirole (2003).

16. Other interpretations of the same structure are also possible; for example, the two pieces of information might be (i) whether moral considerations should enter the decision, and (ii) whether it is moral to perform the action. The important element is a qualifier that modifies ego utility from beliefs about the task itself.

conjectured to bring about an unpleasant mental state called dissonance arousal, which subjects alleviate by changing their opinion about the act they just committed.[17] For the purposes of this paper, the important prediction is the first part (dissonance arousal), and the second (attitude change) is only a proven diagnostic tool to identify the presence of the first.[18] Based on the evidence, dissonance arousal seems to be a low-utility state, with characteristics akin to stressful conditions (Croyle and Cooper 1983; Petty and Cacioppo 1981). In addition, it originates in a negative judgment about the self—if subjects perceive that their actions are harmless, dissonance arousal does not occur.[19] Thus, dissonance arousal is a manifestation of ego utility.

In addition to strong evidence on the existence of ego utility, there is some evidence that—consistent with my assumption about functional form—people are more averse to self-relevant information when they happen to hold high beliefs about themselves than when they happen to hold low beliefs. For example, Stahlberg, Petersen, and Dauenheimer (1999) report that subjects show greater interest in further information after receiving negative (bogus) feedback than after receiving positive feedback. Importantly, this is only true for dimensions along which the subject has uncertain self-knowledge.[20] In another study, Wyer and Frey (1983) show that subjects who had been given negative (bogus) feedback on an IQ test recalled more (positive and negative) information from subsequent articles about IQ tests than did those given positive feedback. Thus, people with a lowered self-esteem seem to search more for information.

These studies also indicate that the current paper's model of ego utility is closer to reality than an alternative specification in which individuals derive utility directly from the signals they receive about themselves (rather than their beliefs). If that was the case, there should not be systematic differences in preference for information depending on whether beliefs are uncertain. In addition, such a model predicts that because they can expect the most favorable news, individuals with high rather than low self-esteem seek out the most information about themselves.

There is also some evidence that people seek out information consistent with their view of themselves, be that view positive or negative (see Fiske and Taylor 1991 for a discussion). This paper abstracts from such "self-consistency" theories to focus on self-enhancement.

17. For example, a subject who wrote a pro-marijuana essay for children feels bad due to the dissonance between her view of herself as a good person and her realization that she committed a bad act. To reduce this dissonance, she "convinces herself" that marijuana use is not so bad after all.

18. See Cooper, Zanna, and Taves (1978) and Zanna and Cooper (1976).

19. For example, if subjects make a counterattitudinal speech on the legalization of marijuana to a non-committed audience, they experience greater attitude change than when the audience is firmly committed in either direction (Nel, Helmreich, and Aronson 1969). For a summary of related findings, see Cooper and Fazio (1984).

20. On the issue of uncertainty, see also Harris and Snyder (1986).

Although my model adds ego utility to the decisionmaker's payoff space, it remains quite classical in that it assumes Bayesian updating, putting a very strong restriction on what the decisionmaker can believe about herself. This contrasts with the views of some psychologists (e.g., Taylor and Brown 1988) that people often have irrationally positive beliefs about themselves. But our ability to reward ourselves by self-delusion is limited (Heider 1958; Tetlock and Levi 1982). The model explores the consequences of such a limitation by making the extreme assumption of Bayesian updating.

More importantly, there is a strong methodological reason for using the Bayesian model. Many psychologists believe that the self-image motivation is the main force behind the biases discussed in this paper (Kunda 1987; and others). Following standard methodology, to isolate the effect of this force I am introducing no other new element into the model.

5.2. Evidence on the Predictions of the Model

The empirical validity of the main prediction of the paper, overconfidence, is beyond question: People hold rather favorable views of their abilities. In a survey by Meyer (1975), more than 90% of employees rated themselves as above the median compared to others in similar jobs and with similar salaries.[21] Svenson (1981) reports that 90% of drivers consider themselves better than the median driver. Biases are found in reports of intelligence (Wells and Sweeney 1986), leadership ability, and memory (Campbell 1986). Similarly, people are more likely to attribute positive than negative outcomes to the self (e.g., Taylor and Brown 1988) and to believe that their first judgments are right (e.g., Lord, Ross, and Lepper 1979).[22]

It is unlikely that people report these views purely for self-presentational purposes. The biases respond little to manipulations about anonymity or whether subjects believe lies can be detected (Greenwald and Breckler 1985).[23] In addition, some studies indicate that subjects act on biased beliefs. For example,

21. The average rating was at the 77th percentile. Only 2 out of about 100 surveyed rated themselves below the median—both putting themselves at the 45th percentile.

22. Unrealistic views are not limited to ability. In general, people seem to be overoptimistic about their prospects. They estimate the likelihood that they will experience a variety of pleasant events (getting a good salary, living past 80, etc.) to be higher than for their peers (Weinstein 1980). At the same time, people think unpleasant events (being in an automobile accident, being a crime victim, having gum problems, etc.) are less likely to happen to them than to others (Perloff and Fetzer 1986).

An exception to these optimistic biases is smoking-related illnesses. People grossly overestimate the negative effects of smoking (Viscusi 1994).

23. With the "bogus pipeline manipulation", most subjects can be made to believe that the lie detector they are connected to works reliably. This manipulation decreases biases somewhat, but the biases do not disappear (Riess et al. 1981).

Camerer and Lovallo (1999) find excess entry in a market game when the payoffs to entering depend on skill, but not when they depend on chance. See also Staw (1974).

There is some evidence for the specific predictions in Section 3.1 on the form of beliefs. Most interesting, beliefs about the self seem to have the property that more confident people are overconfident, while the less confident have unbiased self-knowledge (e.g., Coyne and Gotlib 1983).

5.3. Related Literature

There are two broad categories of non-Bayesian models that predict biased self-views. In one set of models, decision makers directly choose their beliefs, and prefer to become overly optimistic rather than accurate to make themselves feel better (Akerlof and Dickens 1982; Yariv 2002; Brunnermeier and Parker 2004). In other models, agents can acquire biased beliefs by systematically misinterpreting new information (Daniel, Hirshleifer, and Subrahmanyam 1998; Gervais and Odean 2001).[24] This paper shows that biased beliefs can follow purely from ego utility (without biased information processing) and in contrast to the cited papers also studies how the agent manages her self-image by often distorting financial choices relative to her beliefs.

There are also several Bayesian models in which agents acquire biases in certain measures. See Carrillo (1998), Carrillo and Mariotti (2000), and Bénabou and Tirole (2002) in the context of hyperbolic discounting; Zábojnik (2004) in the context of costly learning with type-dependent costs; and Rabin (1995) in the context of moral constraints and self-serving biases. These theories, however, often predict pessimistic rather than optimistic biases. In hyperbolic discounting models, for instance, it might be beneficial to believe that unprotected sex leads very easily to contracting HIV, because future selves' impulsive behavior is kept in control by this "fear." Therefore, the agent stops looking for evidence on HIV when she thinks it is easily transmitted. Although this prediction matches the pessimistic beliefs many people seem to have regarding the contagiousness of HIV, the same logic applies to the dangers of cancer and many other diseases as well, where optimistic biases have been observed. Also, if ability and effort are substitutes in a task, a model similar to Bénabou and Tirole (2002) implies negative views about ability, as negative self-views

24. The incentives to manipulate self-image are similar in flavor to risk-taking incentives generated by career concerns with a discontinuous reward structure. Depending on their current perceptions, managers might be too conservative or too risk-seeking to maximize the probability of being rehired (Zwiebel 1995). Zwiebel does not investigate the consequences of career concerns for biases in self-perception. It is clear, however, that career concerns cannot generate biases in an individual decision-making context, or even in interpersonal settings where information-acquisition decisions are private.

induce future selves to exert greater effort. Finally, as mentioned herein, in addition to generally predicting overconfidence this paper also studies the distinction and interplay between confidence in beliefs and confidence in observed actions.[25]

6. Conclusion

This paper models some implications of ego utility for information gathering and economic decision making. By extending the payoff space to include beliefs about the self, one can incorporate this motivational factor into models with a strictly economic methodology.

The model in this paper, however, is of a rather reduced-form nature. Information gathering is limited to information about one's ability to do well in the financial decision periods; how ability influences success is not explicitly modeled. By modeling the decision problem in greater detail, one can gain insight into more specific modes of self-image protection and enhancement. This is done in my companion paper (Köszegi 2000).

An agenda for future research would include studying the implications of ego utility in economic applications—such as unemployment, warm glow, and public goods, dangerous habits like smoking, and the intergenerational transmission of beliefs—where people's self-image seems to be strongly implicated. Furthermore, it would be interesting to investigate how ego utility affects strategic interactions. On the simplest level, it is clear that through overconfidence, self-image protection, and self-image enhancement, ego utility can act as a commitment device for playing certain strategies with higher probabilities. This, in turn, can alter the nature of a game.[26] On a more subtle level, concern for one's self-image can influence one's view of other players' strategies, indirectly changing the game.[27] Finally, the ego utility of another player can be exploited to elicit certain forms of behavior. Thus, ego utility might be a platform on which to formalize psychological pressure, something that has not been studied with economic tools.

25. Bodner and Prelec (1996) study a (Bayesian) multiple-self model, in which one self manipulates the signals received by the other self in order to increase utility from self-image. They do not address whether this leads to biased beliefs.

26. Consider, for instance, bargaining or "fights": if ego utility makes it more likely that one of the players is going to fight, the other one is more likely to back down, which can be good for the first player even in the instrumental-utility sense.

27. Consider the following model of teenage revolts. The teenager begins with a self-image enhancement motive, and chooses one of two kinds of activities: One in which her parents can help, or one in which they cannot. The parents, who want their child to succeed, then decide whether or not to help, where helping undermines the self-image enhancement motive by making success less satisfying for the teen. If the parents do not care enough about their child's ego utility, they will want to help. Anticipating this, the teenager "revolts" and chooses the activity in which the parents are powerless to interfere.

Appendix A. Notation

In the proofs below, I will use the following notation in addition to that introduced in the text. For any beliefs F and set of signals S, let $F_0^j(F,S)$, $F_1(F,S)$, or $F_2(F,S)$ denote the beliefs derived from applying Bayes' rule to the prior F and the signals S (where the subscripts denote the period in which the posterior beliefs are held, and the superscript j in F_0^j is the number of signals observed). Furthermore, let $U_2(F)$ denote the agent's value function in period 2, when she enters that period with beliefs F. Finally, let Φ and ϕ denote the cumulative distribution and probability density functions of the standard normal, respectively.

Appendix B. Proofs

Proof of Proposition 1. We prove that an optimal strategy exists for any normally distributed initial beliefs. Then, the existence of optimal continuation strategies follows by the redefinition of initial beliefs.

Once the agent stops drawing signals in period 0 and enters period 1 with beliefs F, her optimization problem can be solved by backward induction for each of Assumptions A, B, or C. As in the text, let her maximum utility be $U(F)$. Note that (holding the variance of F constant) $U(F)$ is continuous in μ_F except possibly at zero, so it is measurable in μ_F.

For any positive integer J, consider a "truncated" decision-making problem in which the decision maker can draw at most J signals in period 0. We prove by backward induction that this problem has a solution with measurable value functions $V^{j,J}:\mathbb{R}\to\mathbb{R}$ mapping the mean of current beliefs to expected utility when j signals have been observed, and with measurable optimal strategies $a_0^{j,J}:\mathbb{R}\to\{0,1\}$, where $a_0^{j,J}(\mu)=1$ if the agent draws the next signal when she has mean beliefs μ after j signals. Then, the decision maker's expected utility under the optimal strategy in the truncated problem will be given by $V^{0,J}(\mu_a)$.

Because $V^{J,J}$ is given by U, it is measurable. Now, suppose that for some $j+1\leq J$, $V^{j+1,J}$ is measurable. Then, for an agent holding beliefs F_0^j after having observed j signals, her value function is given by

$$V^{j,J}\left(\mu_{F_0^j}\right)=\max\left\{U(F_0^j),\,E_{F_0^j}[V^{j+1,J}(\mu_{F_0^{j+1}(F_0^j,s_0^{j+1})})]\right\}.$$

Because the maximum of two measurable functions is measurable, $V^{j,J}$ is measurable. Furthermore, an optimal strategy has $a_0^{j,J}(\mu_{F_0^j})=0$ if and only if $V^{j,J}(\mu_{F_0^j})=U(F_0^j)$. Because the set of points on which a measurable function is greater than another measurable function is measurable, $a_0^{j,J}$ is measurable.

Because the agent can choose never to observe the signal $j+1$ even if she has the opportunity to do so, we must have $V^{0,J+1}(\mu_a)\geq V^{0,J}(\mu_a)$. Because

$V^{0,J}(\mu_a)$ is also bounded, it has a limit. Let $V^* = \lim_{J\to\infty} V^{0,J}(\mu_a)$. Clearly, V^* is the supremum of the utility levels the agent can achieve in the original decision problem (with the possibility to observe an infinite number of signals). We complete the proof by showing that the agent can actually achieve this level of utility.

By Tychonoff's theorem, the space of functions from $\mathbb{R}$ to $\{0, 1\}$ is compact in the topology of pointwise convergence. Thus, using the diagonal method, we can select a sequence $J_k \to \infty$ such that a_0^{j,J_k} converges pointwise for each j. Let the limit be a_0^j. Because the pointwise limit of measurable functions is measurable (Aliprantis and Border 1999, Theorem 4.26), a_0^j is measurable. By Lesbegue's dominated convergence theorem, the agent's expected utility if she follows the strategies a_0^j is V^*. □

Proof of Proposition 2. We show that for any beliefs F_0^j satisfying the conditions of the proposition,

$$E_{F_0^j}\big[U(F_0^{j+1}(F_0^j, s_0^{j+1}))\big] \geq U\big(F_0^j\big), \tag{A.1}$$

with strict inequality whenever $\mu_{F_0^j} < 0$. This immediately implies that the agent samples s_0^{j+1} if $\mu_{F_0^j} < 0$. Even if $\mu_{F_0^j} = 0$, we have $\mathrm{Prob}\big(\mu_{F_0^{j+1}(F_0^j, s_0^{j+1})} < 0\big) > 0$, so sampling s_0^{j+1} and then also s_0^{j+2} whenever $\mu_{F_0^{j+1}(F_0^j, s_0^{j+1})} < 0$ dominates stopping after j signals.

Inequality A.1 has been established in the text for Assumption A. For Assumption B, the argument is very similar. Notice that whenever $\mu_{F_0^j} \leq 0$,

$$E_{F_0^j}\big[u(F_1(F_0^j, s_0^{j+1}, s_1))\big] \geq E_{F_0^j}\big[u(F_1(F_0^j, s_1))\big],$$

because two signals are more likely to push mean beliefs above zero than one signal is. Furthermore, the inequality is strict if $\mu_{F_0^j} < 0$. Similarly

$$E_{F_0^j}\big[u(F_2(F_0^j, s_0^{j+1}, s_1, s_2))\big] \geq E_{F_0^j}\big[u(F_2(F_0^j, s_1, s_2))\big],$$

with strict inequality if $\mu_{F_0^j} < 0$. As in the case of Assumption A, the agent's expected instrumental utility does not decrease if she observes an extra signal, because she can ignore this signal in choosing her actions. This completes the proof.

Finally, suppose Assumption C holds. Let $a_i^*(F)$, $i = 1, 2$, denote the optimal strategy in period i when the agent enters that period with beliefs F. We distinguish two cases.

Case 1. $a_1^*(F_0^j) = 0$. Then, we can show by contradiction that $a_2^*(F_0^j) = 0$; if the opposite was the case, it would be strictly better to choose $a_1 = 1$ and $a_2 = 0$

irrespective of s_1. (Choosing the ambitious action first gives the same expected instrumental utility. But it gives strictly higher expected ego utility, because if the one observed signal pushes mean beliefs above zero, the agent can enjoy the extra ego utility for two periods.) Thus,

$$U\big(F_0^j\big) = 2wu\big(F_0^j\big) = 0$$
$$< 2wE_{F_0^j}\big[u(F_0^{j+1}(F_0^j, s_0^{j+1}))\big] \le E_{F_0^j}\big[U(F_0^{j+1}(F_0^j, s_0^{j+1}))\big],$$

because the agent can choose $a_1 = a_2 = 0$ regardless of s_0^{j+1}.

Case 2. $a_1^*(F_0^j) = 1$. Then

$$U\big(F_0^j\big) = E_{F_0^j}\big[wu(F_1(F_0^j, s_1)) + x_1(1, s_1)\big] + E_{F_0^j}\big[U_2(F_1(F_0^j, s_1))\big], \quad \text{(A.2)}$$

Because the mean of F_0^j is negative, $E_{F_0^j}[x_1(1, s_1)] \le 0$. Using that the distributions of s_1 and s_0^{j+1} are the same, the above is less than or equal to

$$E_{F_0^j}\big[wu(F_1(F_0^j, s_0^{j+1})) + x_1(0, s_1)\big] + E_{F_0^j}\big[U_2(F_1(F_0^j, s_0^{j+1}))\big], \quad \text{(A.3)}$$

which is the decision maker's expected utility from drawing s_0^{j+1}, choosing $a_1 = 0$, and then following the optimal continuation strategy. This expected utility is less than or equal to $E_{F_0^j}[U(F_0^{j+1}(F_0^j, s_0^{j+1}))]$. Furthermore, since $E_{F_0^j}[x_1(1, s_1)] < 0$ if $\mu_{F_0^j} < 0$, Expression A.2 is strictly less than Expression A.3 in that case. This completes the proof. □

Proof of Proposition 3. We prove by contradiction; suppose that with beliefs F_0^j, the agent chooses to find out her type with probability 1. This certainly implies that she chooses to observe the next signal. Let F_0^{j+1} be her beliefs after observing s_0^{j+1}. For notational simplicity, let $\mu_0 = \mu_{F_0^{j+1}}$ and denote the variance of F_0^{j+1} by σ_0^2. We show that if μ_0 is sufficiently high, the agent prefers not to follow the strategy of finding out her ability with probability one—a contradiction.

If the agent chooses to find out her type, then under any of the Assumptions A, B, or C, an optimal strategy for her to follow in period $t \in \{1, 2\}$ is to choose $a_t = 1$ if and only if $q > 0$. Call this strategy I. Consider instead strategy II, the strategy of acquiring no further information in period 0, and choosing $a_t = 1$ in both periods. These two strategies lead to the same instrumental utility whenever $q > 0$, but (due to optimal financial choices) strategy I yields higher expected instrumental utility when $q < 0$. For a given q, strategy II yields a per-period expected instrumental payoff of

$$m(q) \equiv 1 - 2\Phi\left(-\frac{q}{\sigma_s}\right),$$

whereas strategy I yields zero. The difference in expected instrumental utility between strategies II and I is thus

$$\frac{2}{\sigma_0}\int_{-\infty}^{0}\phi\left(\frac{q-\mu_0}{\sigma_0}\right)m(q)\,dq.$$

We now look on the ego utility side. The probability that $q < 0$ is $\Phi(-\mu_0/\sigma_0)$. Thus, if the agent chooses strategy I, her ego utility will be zero with this probability. Under strategy II, the same probability is zero if Assumption A is satisfied. Under Assumptions B or C, the agent's ego utility is zero in period 1 if s_1 is sufficiently small to push her mean beliefs below zero. Note that upon observing s_1, the mean of the agent's updated beliefs is

$$\frac{\sigma_s^2}{\sigma_s^2+\sigma_0^2}\mu_0+\frac{\sigma_0^2}{\sigma_s^2+\sigma_0^2}s_1,$$

which is negative if $s_1 \le -(\sigma_s^2/\sigma_0^2)\mu_0$. Because from the perspective of current beliefs, $s_1 \sim N(\mu_0, \sigma_0^2+\sigma_s^2)$, the probability that the agent will have zero ego utility in period 1 is

$$\Phi\left(-\frac{\sqrt{\sigma_0^2+\sigma_s^2}}{\sigma_0^2}\mu_0\right). \tag{A.4}$$

By a similar calculation, the same probability for period 2 is

$$\Phi\left(-\frac{\sqrt{\sigma_0^2+\frac{1}{2}\sigma_s^2}}{\sigma_0^2}\mu_0\right).$$

We now put these calculations together. For all Assumptions A, B, or C, the difference between expected utility from strategy II and expected utility from strategy I is at least

$$2w\Phi\left(-\frac{\mu_0}{\sigma_0}\right)-w\Phi\left(-\frac{\sqrt{\sigma_0^2+\sigma_s^2}}{\sigma_0^2}\mu_0\right)$$
$$-w\Phi\left(-\frac{\sqrt{\sigma_0^2+\frac{1}{2}\sigma_s^2}}{\sigma_0^2}\mu_0\right)+\frac{2}{\sigma_0}\int_{-\infty}^{0}\phi\left(\frac{q-\mu_0}{\sigma_0}\right)m(q)\,dq.$$

We prove that this expression is greater than zero for a sufficiently high μ_0.

Dividing by $\Phi(-\mu_0/\sigma_0)$ gives

$$2w - w\frac{\Phi\left(-\frac{\sqrt{\sigma_0^2+\sigma_s^2}}{\sigma_0^2}\mu_0\right)}{\Phi\left(-\frac{\mu_0}{\sigma_0}\right)} - w\frac{\Phi\left(-\frac{\sqrt{\sigma_0^2+\frac{1}{2}\sigma_s^2}}{\sigma_0^2}\mu_0\right)}{\Phi\left(-\frac{\mu_0}{\sigma_0}\right)}$$
$$+\frac{2}{\sigma_0}\int_{-\infty}^{0}\frac{\phi\left(\frac{q-\mu_0}{\sigma_0}\right)}{\Phi\left(-\frac{\mu_0}{\sigma_0}\right)}m(q)\,dq. \tag{A.5}$$

The first of these four terms is constant in μ_0. We prove that the other three approach zero as $\mu_0 \to \infty$ by exploiting the following property of normal distributions. If $q \sim N(\mu_0, \sigma_0^2)$, let G_{μ_0} denote the distribution of q conditional on $q \leq 0$. As $\mu_0 \to \infty$, G_{μ_0} becomes infinitely right-weighted; that is, for any $\varepsilon, \delta > 0$ there is a M such that if $\mu_0 \geq M$, then $G_{\mu_0}(-\delta) < \varepsilon$. The second term in expression A.5 is

$$wG_{\mu_0}\left(-\left(\frac{\sqrt{\sigma_0^2+\sigma_s^2}}{\sigma_0}-1\right)\mu_0\right),$$

and the third one is

$$wG_{\mu_0}\left(-\left(\frac{\sqrt{\sigma_0^2+\frac{1}{2}\sigma_s^2}}{\sigma_0}-1\right)\mu_0\right),$$

both of which approach zero as $\mu_0 \to \infty$ by the property above. Now we turn to the last term. Notice that the expected instrumental utility $m(q)$ in that term is integrated with respect to G_{μ_0}. Because $m(0) = 0$ and m is continuous, this implies that the expectation of m with respect to G_{μ_0} approaches zero as $\mu_0 \to \infty$. □

Proof of Proposition 4. Let $a_2^*(F_1(F_0^j, s_1))$ denote the optimal period-2 continuation action when the agent observes s_1. If she chooses $a_1 = 0$, she receives no information in period 1. Thus, the strategy $a_1 = 1, a_2 = a_2^*(F_1(F_0^j, s_1))$ is optimal if and only if it is superior to two alternative strategies: $a_1 = 0, a_2 = 1$, and $a_1 = a_2 = 0$. First, we prove that (holding j constant) the difference in expected utility between choosing $a_1 = 1, a_2 = a_2^*(F_1(F_0^j, s_1))$ and choosing $a_1 = 0, a_2 = 1$ strictly increases in $\mu_{F_0^j}$. Notice that

$$E_{F_0^j}\left[wu(F_1(F_0^j, s_1)) + x_1(s_1, 1)\right] = E_{F_0^j}\left[wu(F_2(F_0^j, s_2)) + x_2(s_2, 1)\right].$$

That is, the agent's expected first-period utility with $a_1 = 1$ is equal to her expected second-period utility if $a_1 = 0, a_2 = 1$. Therefore, the difference in expected utility between the two strategies is

$$E_{F_0^j}\big[U_2(F_1(F_0^j, s_1))\big] - E_{F_0^j}\big[wu(F_0^j)\big],$$

the difference between the agent's second-period expected utility when following $a_1 = 1, a_2 = a_2^*(F_1(F_0^j, s_1))$ and her expected first-period utility when following $a_1 = 0, a_2 = 1$. The first of these terms is obviously strictly increasing in $\mu_{F_0^j}$, while the second one is constant.

Next, we prove that the difference in expected utility between choosing $a_1 = 1, a_2 = a_2^*(F_1(F_0^j, s_1))$ and choosing $a_1 = a_2 = 0$ also strictly increases in $\mu_{F_0^j}$. This is easy: The expected utility from following the first strategy strictly increases in $\mu_{F_0^j}$, while the expected utility from following the second one is constant.

Now notice that the expected utility from following $a_1 = 1, a_2 = a_2^*(F_1(F_0^j, s_1))$ approaches $2(1 + w)$ as $\mu_{F_0^j} \to \infty$, because the probability that the agent will receive $x_1 = x_2 = 1$, and the probability that she has positive mean beliefs in both periods, approach one. Furthermore, if $a_1 = 0$, the agent's expected utility is bounded by $2w + 1$, because in that case $x_1 = 0$. Thus, for a sufficiently high $\mu_{F_0^j}$, following $a_1 = 1, a_2 = a_2^*(F_1(F_0^j, s_1))$ is strictly better than either of the above strategies. As a result, there is a unique $\mu' \geq 0$ such that choosing $a_1 = 1, a_2 = a_2^*(F_1(F_0^j, s_1))$ dominates $a_1 = 0, a_2 = 1$ if and only if $\mu_{F_0^j} > \mu'$; and there is a unique $\mu'' \geq 0$ such that choosing $a_1 = 1, a_2 = a_2^*(F_1(F_0^j, s_1))$ dominates $a_1 = a_2 = 0$ if and only if $\mu_{F_0^j} > \mu''$. The agent chooses $a_1 = 1$ if and only if $\mu_{F_0^j} > \mu_1^j \equiv \max\{\mu', \mu''\}$. □

Proof of Proposition 5.

1. Take w and w', with $w > w'$. Let U_{ego} and U_x be the agent's expected ego utility ($E[u(F_1) + u(F_2)]$) and instrumental utility ($E[x_1 + x_2]$), respectively, in an optimal strategy with weight w on ego utility. Define U'_{ego} and U'_x similarly, except with weight w' on ego utility. By revealed preference,

$$wU_{\text{ego}} + U_x \geq wU'_{\text{ego}} + U'_x$$
$$w'U'_{\text{ego}} + U'_x \geq w'U_{\text{ego}} + U_x.$$

Adding and rearranging gives

$$(w - w')(U_{\text{ego}} - U'_{\text{ego}}) \geq 0,$$

which implies that $(U_{\text{ego}} - U'_{\text{ego}}) \geq 0$.

2. First, we prove that if $\mu_{F_0^j} \leq 0$ for the agent's current beliefs F_0^j in period 0, then she must attach positive probability to the possibility that she will never

have mean beliefs strictly above zero. Suppose by contradiction that she reaches positive mean beliefs with probability one. Let F denote her beliefs the first time she reaches positive mean beliefs (F is of course a random variable). Let f_0^j and f denote the probability density functions of F_0^j and F, respectively. By the law of iterated expectations, for any $q \in \mathbb{R}$,

$$f_0^j(q) = E_{F_0^j} f(q).$$

Because F has a positive mean and a lower variance than F_0^j, there is a uniform bound $\bar{q}$ such that if $q < \bar{q}$, then $f_0^j(q) > f(q)$ for all F. But this contradicts the above equality.

Denote by $p^j(\mu_{F_0^j})$ the minimum probability (over the set of continuation strategies) that if the agent has beliefs F_0^j after drawing j signals, she will never have positive beliefs again. We have $p^j(\mu_{F_0^j}) = 0$ if $\mu_{F_0^j} > 0$, but as we have established above, $p^j(\mu_{F_0^j}) > 0$ for $\mu_{F_0^j} \leq 0$. Now for any F_0^j with $\mu_{F_0^j} > 0$, the probability that the agent ends up with mean beliefs less than zero if she draws another signal is at least $E_{F_0^j}[p^{j+1}(\mu_{F_0^{j+1}(F_0^j, s_0^{j+1})})]$. Because this expectation is strictly greater than zero and continuous in $\mu_{F_0^j}$, we can choose non-negative $\mu^j(w)$ such that (i) for any $w > 0$, if $\mu_{F_0^j} < \mu^j(w)$, then the probability that the agent's mean beliefs will drop below zero and will stay there if she draws another signal is at least $1/w$; and (ii) for all j, $\mu^j(w) \to \infty$ as $w \to \infty$.

Notice that if $0 < \mu_{F_0^j} < \mu^j(w)$, then the agent prefers to stop drawing signals and choose the unambitious option in both periods. Relative to this strategy, continuing drawing signals or choosing the ambitious option in any period costs at least $1/w$ in ego utility per period, and has an instrumental benefit strictly less than 1 per period.

To complete the proof, note that the probability that the agent ends up with negative mean beliefs or stops acquiring information with beliefs F_0^j satisfying $0 < \mu_{F_0^j} < \mu^j(w)$—in both of which cases she chooses the unambitious option in both periods—approaches one as $w \to \infty$. $\square$

Proof of Proposition 6. Suppose the agent holds beliefs $F_1 \sim N(\mu_1, \sigma_1^2)$, and suppose first that $\mu_1 > 0$. If the agent chooses the unambitious option, she receives utility w in period 2. Because the subjective distribution of s is $N(\mu_1, \sigma_1^2 + \sigma_s^2)$, the probability that $s_1 \leq 0$ is $\Phi(-\mu_1/\sqrt{\sigma_1^2 + \sigma_s^2})$. Combining this with the same calculation as in the third paragraph of the proof of Proposition 3, if the agent chooses the ambitious option, she receives expected utility

$$1 - 2\Phi\left(\frac{-\mu_1}{\sqrt{\sigma_1^2 + \sigma_s^2}}\right) + w\left(1 - \Phi\left(-\frac{\sqrt{\sigma_1^2 + \sigma_s^2}}{\sigma_1^2}\mu_1\right)\right).$$

Clearly, the difference in expected utility between the ambitious and unambitious options is strictly increasing and continuous in μ_1, negative for $\mu_1 = 0$, and positive for a sufficiently high μ_1. Thus, there is a unique $\mu_2^j > 0$ such that the agent chooses the unambitious option for $\mu_1 \leq \mu_2^j$.

A symmetric proof gives the result for $\mu_1 \leq 0$. □

Proof of Proposition 7.

1. By revealed preference, for any beliefs F and $a_t \in \mathbb{R}$ in period $t \in \{1, 2\}$,

$$E_F[x_t(s_t, a_t)] \leq E_F[E[x_t(s_t, a^*(\delta_q))|q]].$$

That is, the agent can achieve at least as high expected instrumental utility if she finds out her type and takes the optimal action for her type, than if she follows any other strategy.

Now suppose by contradiction that with the optimal strategy,

$$\frac{E[a^*(F_1) + a^*(F_2)]}{2} < E[a^*(\delta_q)].$$

Thus, if the agent finds out her type, she achieves strictly higher expected ego utility than with her optimal strategy. Given these statements, this implies that she achieves strictly higher overall expected utility by finding out her type and taking the optimal action for her type; thus, we have a contradiction.

2. Let F_0 denote the agent's beliefs at the end of period 0 (which is a random variable). Because the agent receives no information in periods 1 and 2, we have $F_0 = F_1 = F_2$. Thus, by Part 1, $E[a^*(F_0)] \geq E[a^*(\delta_q)]$. Because the agent receives no information in period 1 and her choice of action does not affect her beliefs, she chooses the financially optimal action in each period (given her beliefs). Therefore, $a_1 = a_2 = a^*(F_0)$ at almost every belief F_0 reached, completing the proof.

3. We prove the statement for any F_0 where $a^*(F_0)$ is single-valued. Suppose by contradiction that $a_1 > a^*(F_0)$. Consider instead the following strategy. The agent first draws

$$n \equiv \frac{\sigma_s^2}{\sigma_{a_1}^2} - \frac{\sigma_s^2}{\sigma_{a^*(F_0)}^2}$$

extra signals in period 0. Then, in period 1, she chooses action $a^*(F_0)$. This strategy yields strictly higher expected instrumental utility in period 1. Further, n was chosen so that drawing these signals and choosing $a^*(F_0)$ induce the same distribution of end-of-period-1 beliefs F_1 as not drawing them and choosing a_1.[28] Thus, the alternative strategy yields the same expected ego utility in period 1 and

28. To see this, note that drawing the extra signals and choosing $a^*(F_0)$ increases the precision of beliefs by $1/(\sigma_{a^*(F_0)}^2) + n/\sigma_s^2 = 1/\sigma_{a_1}^2$, which is exactly the precision of the signal observed after choosing a_1.

the same expected continuation value. Combined with the fact that it has strictly higher expected instrumental utility in period 1, this contradicts that a_1 is part of an optimal strategy.

That in the second period the agent does not necessarily choose a lower action than instrumentally optimal is demonstrated by the binary-choice model in the text, which is easily adapted to this framework by assuming that any action $a \notin \{0, 1\}$ yields a payoff less than -1 with probability one. □

Proof of Proposition 8. Let U denote the agent's value function in period 0 for the information structure $k_1, \ldots, k_{i-1}, k_i, k_{i+1}, \ldots$, where U maps possible beliefs into expected utility. Define U' similarly for information structure $k_1, \ldots, k_{i-1}, k_i', k_i'', k_{i+1}, \ldots$. For any beliefs F^{i-1} after the agent has observed $i-1$ groups of signals, $U'(F^{i-1}) \geq U(F^{i-1})$, because the agent can replicate the optimal strategy under the former information structure with the latter information structure. (Whenever she chooses to observe the next k_i signals under the former information structure, she also observes the next k_i' and k_i'' signals under the latter, and chooses the same continuation strategy.) By backward induction, the same is true for any earlier beliefs. □

References

Akerlof, George and William T. Dickens (1982). "The Economic Consequences of Cognitive Dissonance." *American Economic Review*, 72(3), 307–319.

Aliprantis, Charalambos D., and Kim C. Border (1999). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 2nd ed. Springer-Verlag.

Baker, George P., Michael C. Jensen, and Kevin J. Murphy (1988). "Compensation and Incentives: Practice vs. Theory." *Journal of Finance*, 43(3), 593–616.

Baumeister, Roy F. (1998). "The Self." In *The Handbook of Social Psychology*, vol. 1, edited by Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey. McGraw-Hill.

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

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

Bodner, Ronit, and Drazen Prelec (1996). "The Emergence of Private Rules in a Self-signaling Model." *International Journal of Psychology*, 31, 3652–3653.

Brandenburger, Adam M., and Barry J. Nalebuff (1997). *Co-opetition: 1. A Revolutionary Mindset that Redefines Competition and Cooperation; 2. The Game Theory Strategy that's Changing the Game of Business*. Currency Doubleday.

Brunnermeier, Markus, and Jonathan Parker (2004). "Optimal Expectations." NBER Working Paper No. W10707.

Camerer, Colin, and Dan Lovallo (1999). "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review*, 89(1), 306–318.

Campbell, Jennifer D. (1986). "Similarity and Uniqueness: The Effects of Attribute Type, Relevance, and Individual Differences in Self-esteem and Depression." *Journal of Personality and Social Psychology*, 50(2), 281–294.

Caplin, Andrew, and John Leahy (2001). "Psychological Expected Utility and Anticipatory Feelings." *Quarterly Journal of Economics*, 116(1), 55–80.

Carrillo, Juan (1998). "Self-control, Moderate Consumption, and Craving." CEPR Discussion Paper No. 2017.
Carrillo, Juan, and Thomas Mariotti (2000). "Strategic Ignorance as a Self-disciplining Device." *Review of Economic Studies*, 67, 529–544.
Cooper, Joel, and Russell H. Fazio (1984). "A New Look at Dissonance Theory." In *Advances in Experimental Social Psychology*, vol. 17, edited by Leonard Berkowitz. Academic Press.
Cooper, Joel, Mark P. Zanna, and Peter A. Taves (1978). "Arousal as a Necessary Condition for Attitude Change Following Induced Compliance." *Journal of Personality and Social Psychology*, 36(10), 1101–1106.
Coyne, James C., and Ian H. Gotlib (1983). "The Role of Cognition in Depression: A Critical Appraisal." *Psychological Bulletin*, 94, 472–505.
Croyle, Robert T., and Joel Cooper (1983). "Dissonance Arousal: Physiological Evidence." *Journal of Personality and Social Psychology*, 45, 782–791.
Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam (1998). "Investor Psychology and Security Market Under- and Overreactions." *Journal of Finance*, 53(3), 1839–1885.
Deci, Edward (1972). "The Effects of Contingent and Non-contingent Rewards and Controls on Intrinsic Motivation." *Organizational Behavior and Human Performance*, 8, 217–229.
Festinger, Leon (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.
Fiske, Susan T., and Shelley E. Taylor (1991). *Social Cognition*, 2nd ed. McGraw-Hill.
Gervais, Simon, and Terrance Odean (2001). "Learning To Be Overconfident." *Review of Financial Studies*, 14(1), 1–27.
Greenwald, Anthony G. (1988). "The Totalitarian Ego: Fabrication and Revision of Personal History." *American Psychologist*, 35(7), 603–618.
Greenwald, Anthony G., and Steven J. Breckler (1985). "To Whom Is the Self Presented?" In *The Self and Social Life*, edited by B. Schlenker. McGraw-Hill.
Harris, Robert N., and C. R. Snyder (1986). "The Role of Uncertain Self-esteem in Self-handicapping." *Journal of Personality and Social Psychology*, 51(2), 451–458.
Heider, Fritz (1958). *The Psychology of Interpersonal Relations*. Wiley.
Köszegi, Botond (2000). "Ego Utility and Information Acquisition." Working Paper, University of California, Berkeley.
Kunda, Ziva (1987). "Motivated Inference: Self-serving Generation and Evaluation of Causal Theories." *Journal of Personality and Social Psychology*, 53(4), 636–647.
Lepper, Mark R., David Greene, and Richard E. Nisbett (1973). "Undermining Children's Intrinsic Interest with Extrinsic Reward: A Test of the 'Overjustification' Hypothesis." *Journal of Personality and Social Psychology*, 28(1), 129–137.
Lord, Charles G., Lee Ross, and Mark R. Lepper (1979). "Biased Assimilation and Attitude Polarization: The Effect of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology*, 37(11), 2098–2109.
Meyer, Herbert H. (1975). "The Pay-for-Performance Dilemma." *Organizational Dynamics*, 3(3), 39–65.
Milkovich, George T., and Jerry M. Newman (1987). *Compensation*, 2nd ed. Business, Publications, Plano, TX.
Murray, Sandra L., and John G. Holmes (1994). "Storytelling in Close Relationships: The Construction of Confidence." *Personality and Social Psychology Bulletin*, 20, 650–663.
Nel, Elizabeth, Robert Helmreich, and Elliot Aronson (1969). "Opinion Change in the Advocate as a Function of the Persuasibility of his Audience: A Clarification of the Meaning of Dissonance." *Journal of Personality and Social Psychology*, 12, 117–124.
Perloff, Linda S., and Barbara K. Fetzer (1986). "Self-other Judgments and the Perceived Vulnerability of Victimization." *Journal of Personality and Social Psychology*, 50, 502–510.
Petty, Richard E., and John T. Cacioppo (1981). *Attitudes and Persuasion: Classic and Contemporary Approaches*. Wm. C. Brown.

Rabin, Matthew (1995). "Moral Preferences, Moral Constraints, and Self-serving Biases." Berkeley Department of Economics Working Paper No. 95–241.

Riess, Marc, Paul Rosenfeld, Valerie Melburg, and James T. Tedeschi (1981). "Self-serving Attributions: Biased Private Perceptions and Distorted Public Descriptions." *Journal of Personality and Social Psychology*, 41(2), 224–231.

Stahlberg, Dagmar, Lars-Eric Petersen, and Dirk Dauenheimer (1999). "Preferences for and Evaluation of Self-relevant Information Depending on the Elaboration of the Self-schemata Involved." *European Journal of Social Psychology*, 29(4), 489–502.

Staw, Barry M. (1974). "Attitudinal and Behavioral Consequences of Changing a Major Organizational Reward: A Natural Field Experiment." *Journal of Personality and Social Psychology*, 29, 742–751.

Svenson, Ola (1981). "Are We Less Risky and More Skillful Than our Fellow Drivers?" *Acta Psychologica*, 18, 473–474.

Taylor, Shelley E., and Jonathan D. Brown (1988). "Illusion and Well-being: A Social Psychological Perspective on Mental Health." *Psychological Bulletin*, 103(2), 193–210.

Tetlock, Philip E., and Ariel Levi (1982). "Attribution Bias: On the Inconclusiveness of the Cognition-motivation Debate." *Journal of Experimental Social Psychology*, 18, 68–88.

Viscusi, W. Kip (1994). "Cigarette Taxation and the Social Consequences of Smoking." NBER Working Paper No. W4891.

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

Wells, L. Edward, and Paul D. Sweeney (1986). "A Test of Three Models of Bias in Self-Assessment." *Social Psychology Quarterly*, 49, 1–10.

Wyer, Robert S., and Dieter Frey (1983). "The Effects of Feedback About Self and Others on the Recall and Judgments of Feedback-relevant Information." *Journal of Experimental Social Psychology*, 19(6), 540–559.

Yariv, Leeat (2002). "I'll See it When I Believe It: A Simple Model of Cognitive Consistency." Cowles Foundation Discussion Paper No. 1352.

Zábojnik, Ján (2004). "A Model of Rational Bias in Self-Assessments." *Economic Theory*, 23(2), 259–282.

Zanna, Mark P., and Joel Cooper (1976). "Dissonance and the Attribution Process." In *New Directions in Attribution Research*, edited by John H. Harvey, William John Ickes, and Robert F. Kidd. Erlbaum.

Zwiebel, Jeffrey (1995). "Corporate Conservatism and Relative Compensation." *Journal of Political Economy*, 103(1), 1–25.