Fragile Self-Esteem*

Botond Kőszegi
Central European University

George Loewenstein
Carnegie Mellon University

Takeshi Murooka
Osaka University

June 24, 2019

Abstract

We develop a model in which a person's self-esteem is determined by sampling from her store of ego-relevant memories in a fashion that in turn depends on her self-esteem. This feedback mechanism can create multiple "self-esteem personal equilibria," making self-esteem fragile — subject to swings that are out of proportion to the objective evidentiary value of new information. Self-esteem is especially likely to be fragile, as well as unrealistic in either the positive or the negative direction, if self-esteem is an important ingredient of overall utility. We show that fragile self-esteem has diverse consequences, such as producing a negative relationship between effort and incentives, and various forms of information-avoidance, with consequences for domains such as education and job search.

*Kőszegi thanks the European Research Council for financial support under Grant #788918. Loewenstein thanks Jacob Rothschild for helpful bibliographical research. Murooka thanks financial support from JSPS KAKENHI (JP16K21740).

Contents

1 Introduction

"Insecurity," as psychologists use the term, refers to a sense of vulnerability or instability in an individual's self-image or ego (Maslow, 1942). Although viewed by psychologists as a virtually universal characteristic, insecurity, like other psychological traits, varies in extremity across individuals. The psychology literature has also drawn a distinction between explicit self-esteem, which is the level of self-esteem that an individual is consciously aware of (and likely communicates to others), and implicit self-esteem, which is a view of oneself, potentially experienced at an unconscious level, that often encompasses self-doubts (Farnham et al., 1999).

The most important manifestation of insecurity is the phenomenon of high-but-fragile self-esteem, which involves "positive self-feelings that are vulnerable to challenge and that require continual promotion and protection" (Kernis et al., 2005). High but fragile self-esteem takes an especially extreme form in, and is one of the key identifying features of, narcissistic personality disorder (Morf and Rhodewalt, 2001). Insecurity, viewed in terms of the explicit/implicit dichotomy, is characterized by a combination of high explicit and low implicit self-esteem (e.g., Hetts et al., 1999, Koole et al., 2001).

Beyond identifying it as a common trait, the psychology literature has documented a wide range of consequences of insecurity. In a paper reviewing the literature on high but fragile self-esteem, Kernis (2001) reports that "among high self-esteem individuals, the more unstable their self-esteem," (a) the greater their anger and hostility proneness; (b) the more severe their depressive symptoms, and the lower they score on various dimensions of psychological well-being; (c) the more defensively they react to perceived transgressions; (d) the more they engage in self-handicapping (i.e., introducing obstacles to their own success); and (e) the more boastful they are following successes. Less surprisingly, and in a pattern opposite to that of self-handicapping, insecurity can result in extremes of motivation and effort aimed at quelling self-doubts, as exemplified by the phenomenon of workaholism.

The purpose of this paper is to formally model insecurity – "fragile self-esteem" — in a way that takes seriously its underlying psychology, and to trace out its economic implications. In the framework presented in Section 2, a person has a store of self-relevant memories that determine how

she should objectively assess herself. Typically, however, she does not have access to an unbiased sample from these memories. The key assumption of the model is that the memories that come to the agent's mind at any point in time depend, in a self-reinforcing fashion, on his current level of self-esteem. People experiencing high immediate self-esteem feel good about themselves, and due to mood-congruent memory, are more likely to recall successful accomplishments. By the same token, those experiencing low immediate self-esteem feel bad about themselves, and are therefore more likely to call to mind failures and social blunders. This dependence of accessible information on self-esteem, and the reverse and more obvious dependence of self-esteem on accessible ego-relevant information, lead to the possibility of multiple *self-esteem personal equilibria* (SEPs) — states of self-esteem that give rise to memories that sustain them. To the extent that a person experiences occasional ego-shocks that restart his sampling from memory, therefore, he can be vulnerable to wild changes in his level of explicit self-esteem.

We establish some basic properties of self-esteem in Section 3. Since a person's beliefs about his ability, talent, or skill — his self-esteem — are crucial ingredients in many economic decisions, from a classical perspective one might expect that most individuals hold beliefs about themselves that are realistic given the objective evidence available to them. Our model, however, follows a completely different logic. When self-esteem is important to how the person feels about himself and therefore greatly influences the memories he calls up, his self-esteem is fragile, ending up unrealistically high or unrealistically low, but never accurate. This prediction connects our model to two hitherto separate literatures. Research by both economists and psychologists has documented (and economic theory has modeled reasons for) the phenomenon that many individuals hold unrealistically high, "overconfident," beliefs about themselves (Johnson and Fowler, 2011). Less well known to economists, research by psychologists has also documented the prevalence of the "impostor syndrome," whereby people view themselves as less capable than an objective evaluation of their success might suggest (Ferrari and Thompson, 2006). Tellingly, both overconfidence and the impostor syndrome often occur in domains that are important to a person's life. To go further, we show that in the limit when self-esteem is extremely important, it is essentially independent of objective evidence. Hence, the level of a person's self-esteem does not convey reliable information

about his true ability or talent. Yet even then, the fragility of his self-esteem, which we define as the smallest shock that can induce a shift to another SEP, is informative about his ability: e.g., the lower is his ability, the more fragile a high SEP, if it is achieved, will be.

Although self-esteem — its level as well as its potential fluctuations — is interesting and important for welfare in itself, a wider range of implications come into play when the agent can take actions related to his self-esteem. In Section 4, we consider the implications of our model when the agent takes ability-dependent actions. We focus on situations in which the effort the agent puts into an activity (e.g., studying or looking for a job) is increasing in his perceived ability, either due to the classical reason that effort and ability are complements, or due to the more psychological reason that the cost of effort is decreasing in how he feels about himself. We then show that a person with fragile self-esteem who is in a low SEP might react to an increase in incentives by decreasing effort. An increase in incentives increases the importance of self-esteem, and because this makes good memories more emotionally distant, it can lower the agent's self-esteem, outweighing the effect of the incentives. In contrast, if the agent is in a high SEP, he always reacts to an increase in incentives by increasing effort.

In Section 5, we consider not how self-esteem affects the agent's actions, but how he can change his behavior to influence his self-esteem. In particular, because changes in salient momentary outcomes can lead to radical changes in self-perception, people may be extremely motivated to take actions and expend effort when their ego is at stake. One way for the agent to do so is to influence outcomes that affect how he feels about himself at the moment. Our theory predicts that if the agent is in a high SEP, then such effort is decreasing in his ability, formalizing the intuition that it is especially low-ability individuals who are desperate to maintain a high ego. Another way for the agent to insulate himself from a drop in self-esteem is to avoid mood-changing shocks in the first place. This provides one possible interpretation of self-handicapping, a widely observed pattern of behavior whereby people who are insecure about their own abilities introduce obstacles to their own success — e.g., by consuming alcohol or selecting impossibly difficult goals — so that if they do in fact fail, they can blame their failure on the obstacles rather than their own deficiencies (Higgins et al., 2013). Psychological accounts conceptualize self-handicapping as

information avoidance (Urdan and Midgley, 2001), which is consistent with many economic models capturing an information-avoidance motive (Carrillo and Mariotti, 2000, Caplin and Eliaz, 2003, Kőszegi, 2003, Oster et al., 2013, Bénabou and Tirole, 2016, Golman et al., 2017); but for a person with plenty of information about himself, one new talk or one new test result should only have a minute effect on self-perception, so that it is only our model that can predict a calibrationally non-trivial motive for self-handicapping in typical situations.

In Section 6, we show that our theoretical framework can help to make sense of a range of phenomena of interest to economists. A common theme in our applications is that individuals with low self-esteem put in too little effort into improving their outcomes, while individuals with high self-esteem put in too much, but this effort is not necessarily of a productive type. For instance, consistent with much evidence of underinvestment in these domains, a discouraged student may study too little, and an unemployed person may spend too little time searching for a job. Conversely, an unemployed individual who has, against the odds, maintained high self-esteem, and a high-self-esteem student, may work too hard. But if either of these individuals is afraid of setbacks that could destroy their high self-esteem, they may spend their effort on insufficiently ambitious tasks, for instance endlessly searching for the appropriate job instead of risking rejections by sending applications. And consistent with the phenomenon of workaholism, a confident employee may put in unreasonably high amounts of effort, even at a severe cost to other, possibly non-ego-related, aspects of life.

Our framework also says that a person who believes that innate ability is the main determinant of achievement is at risk of an emotional collapse after a setback, whereas a person who believes that ability can be developed through effort is not. This provides a new perspective on one of the most prominent set of findings in recent research on education, the effect of a fixed mindset as compared with a growth mindset.

In a final application, we examine the consequences of fragile self-esteem for aggression, a topic that has received the lion's share of attention in the psychology literature dealing with insecurity. Drawing on insights from Roy Baumeister, we argue that people with high but fragile self-esteem are in a precarious situation ego-wise, and will lash out, often at risk to themselves, when doing so

can protect them from what might otherwise be a precipitous drop in self-esteem.

In addition to integrating insights from psychology, our paper builds on recent work by economists that has recognized the importance of ego or self-worth (Brennan and Pettit, 2004), as well as people's uncertainty about their own self-worth, as important motives driving human behavior (e.g., Bénabou and Tirole, 2006, Kőszegi, 2006, Dana et al., 2007, Grossman and van der Weele, 2017). Uncertainty about one's own self-worth is a key enabling condition for people to hold, and to recognize the potential to hold, different views at different points in time. The model we propose also builds on theoretical work by economists that examines implications of biased recall of memories (Bernheim and Thomadsen, 2005, Bordalo et al., 2017). Finally, our paper applies the notion of personal equilibrium developed for utility from beliefs (Kőszegi, 2010), and follows previous papers that emphasize the possibility of multiple personal equilibria in other domains (Kőszegi and Rabin, 2006, Spiegler, 2016).

2 Framework

In this section, we formulate a model of self-esteem determination based on the interaction of memory, feelings about oneself, and self-esteem. The crucial assumption of our model is that the memories a person accesses at any time are influenced by how he feels about himself — or his mood for short. Such differences in retrieval, termed "mood-congruent memory," have been observed in research by psychologists documenting differences in recall by individuals who are or are not subject to mood-related clinical conditions such as depression (Watkins et al., 1992, 1996), within the same individual when in naturally occurring happy or sad states (Mayer et al., 1995), and, perhaps most convincingly, in response to experimentally induced moods (Matt et al., 1992). In one study, positive or negative moods induced via exposure to happy or sad music led to substantial differences in recall of positive and negative life events. Especially relevant to our focus on self-esteem, a large number of studies have shown that individuals in negative moods are more likely than those in positive moods to recall episodes of failure or low task performance relative to episodes of success

and/or high task performance.[1] Other research traces individual proclivities toward depression in part to negative memory biases (Beck, 1979, Solomon, 2014).

Formally, the agent's perception of his ability is derived from sampling his *evidentiary base* $\{s_1, \cdots, s_n\}$, which is composed of facts $s_i \in \mathbb{R}_+$. We think of the evidentiary base as containing all information relevant for arriving at an assessment of ability, including all facts or experiences determining the agent's prior. Consistent with this perspective, we assume that a realistic assessment of ability corresponds to the simple average $a = E[s_i]$. We define the person's actual self-esteem as his perceived ability $\tilde{a}$.

Self-esteem is an important determinant of mood, but not the only determinant. We designate the importance of self-esteem as a determinant of mood with the variable $k > 0$. Normalizing the agent's minimum mood to zero, his mood m is his perceived ability weighted by its importance: $m = k\tilde{a}$. Similarly, $S_i = ks_i$ is the emotional implication of fact s_i; this is how the agent would feel if he thought s_i was his ability. Intuitively, if ability is more important for the agent, then his mood is more sensitive to how positively he perceives his ability, and each fact has a greater emotional impact on him.

There may be both psychological and economic determinants of k. First, individuals may differ in how much they care about ability for personality reasons; some people are simply more 'ego-driven' than others. At the extreme, although by no means alone in caring about ego, are narcissists who are characterized by a "preoccupation with building, buttressing, and defending... [their] desired self" (Morf and Rhodewalt, 2001). Second, it is natural to assume that economic stakes also affect k: if self-confidence, or ego, is more important in determining a person's future outcomes, then he is prone to care more about his perceived ability.

The crucial assumption of our model is that the agent's perceived ability at any point in time depends on what facts he recalls, which in turn depends on his mood. Formally, if the agent's mood is m, then he recalls fact s_i with probability $g(S_i - m)/\left[\sum_{i'} g(S_{i'} - m)\right]$, where $g(\cdot)$ is a

[1] See Blaney (1986) for an extensive review. The phenomenon of mood-congruent memory is closely related to the "mood-congruent judgment" effect (Mayer et al., 1992) whereby mood-congruent information is judged to be more valid and relevant than mood-incongruent information. From a theoretical perspective, the notion that a person in bad mood is more likely to recall failures, and the notion that he judges failures more relevant, are equivalent, as both increase the weight of failure in his beliefs.

three times differentiable positive-valued function that is single-peaked at and symmetric around zero. Intuitively, the further is the emotional implication of a fact S_i from the agent's mood m, the less likely he is to recall the fact. Let $E[s_i|\tilde{a}] = [\sum_i g(k(s_i - \tilde{a}))s_i]/[\sum_i g(k(s_i - \tilde{a}))]$ be the average memory that the agent recalls when he has self-esteem $\tilde{a}$. Self-esteem $\tilde{a}$ is then determined according to:

Definition 1 (Self-Esteem Determination)**.** A level of self-esteem $\tilde{a}$ is a self-esteem personal equilibrium (SEP) if it is a locally stable solution to the equation $\tilde{a} = E[s_i|\tilde{a}]$.[2]

Although we do not formally specify and analyze a stochastic recall process that determines self-esteem, our definition of SEP — as well as our definition of fragility below — are broadly motivated by the following idea. The agent's momentary self-esteem fully determines his mood, which determines the probabilities with which he recalls facts. These in turn affect his new self-esteem and thereby his mood. Through this stochastic recall process, the agent converges to a stable self-esteem $\tilde{a}$. If the agent has experienced stable self-esteem for an extended period, then the average memory evoked from the evidentiary base, and hence his perceived ability, must be $\tilde{a} = E[s_i|\tilde{a}]$: a stable level of self-esteem must on average evoke memories that justify such self-esteem. And for his self-esteem not to drift away from the fixed point in response to random shocks, the fixed point must be locally stable. Although applied in a different context, these motivations are very similar to those in Kőszegi and Rabin (2006) and Kőszegi (2010).

While our framework allows for arbitrary finite databases, for analytical simplicity we derive predictions from a simplified situation in which the evidentiary base is binary. An element of the evidentiary base is $s_i \in \{0, 1\}$, and the proportion of $s_i = 1$ is $a \in [0, 1]$. Our assumptions on $g(\cdot)$ imply that a SEP exists for almost any a:

Fact 1 (Existence)**.** *Given $g(\cdot)$ and $k > 0$, a SEP exists for almost any a.*

In the proof of Fact 1, we also show that a solution to the equation $\tilde{a} = E[s_i|\tilde{a}]$ is locally unique for almost any a (or almost any database). Based on this, we assume in the rest of the paper that any solution to the equation $\tilde{a} = E[s_i|\tilde{a}]$ is locally unique.

[2] The notion of local stability we require is the conventional one: that there is a $\delta > 0$ such that $\tilde{a}' < E[s_i|\tilde{a}']$ for $\tilde{a} - \delta < \tilde{a}' < \tilde{a}$ and $\tilde{a}' > E[s_i|\tilde{a}']$ for $\tilde{a} < \tilde{a}' < \tilde{a} + \delta$.

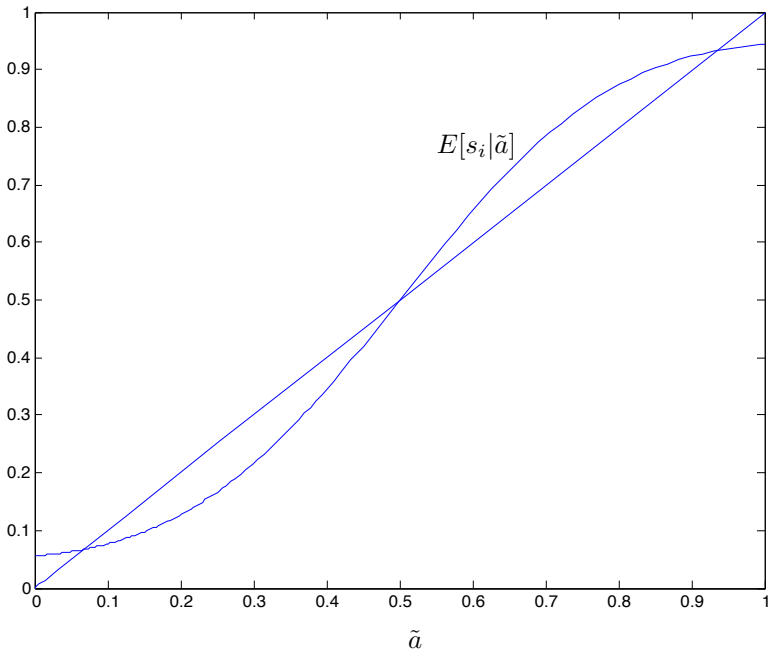

Figure 1: An Example of Multiple SEP

The essential feature of our framework, however, is that there may be multiple SEPs. To illustrate such multiplicity, we introduce a parametric example that we will use repeatedly in the paper:

Example 1. Let $g(x) = 1/(1 + x^2)$.

Figure 1 illustrates SEP determination in this example when actual ability is $a = 1/2$ and $k = 4$. $E[s_i|\tilde{a}]$ is the average fact that the agent recalls when his self-esteem is $\tilde{a}$. By definition, the individual's SEPs then correspond to points where $E[s_i|\tilde{a}]$ intersects the 45-degree line from above. As is apparent in the figure, one intersection is at the true ability $\tilde{a} = 1/2$, but it is not from above (i.e., this solution is not locally stable). The other two intersections, however, are from above (i.e., they are locally stable) and therefore correspond to SEPs. Solving numerically, these are at self-esteem levels $\tilde{a} = 0.067$ and $\tilde{a} = 0.933$. Intuitively, if the agent's self-esteem is low, he does not recall good facts about himself, confirming his low self-esteem; and if his self-esteem is high, he recalls mostly good facts about himself, again confirming his self-esteem.

For some of the insights in the paper, we do not need to specify which SEP the agent ends up at. For some points, however, this is crucial. We make a specific assumption motivated by the recall process that we used to motivate our definition of SEP. Suppose that the agent starts off at initial self-view $\tilde{a}_0$, or, equivalently, his initial feeling about himself is given by $k\tilde{a}_0$. Then, his self-esteem converges to a SEP according to the following:

Definition 2. Given seed self-view $\tilde{a}_0$, the $\tilde{a}_0$-SEP is given by (i) the lowest SEP level $\tilde{a} \geq \tilde{a}_0$ if $E[s_i|\tilde{a}_0] \geq \tilde{a}_0$; and (ii) the highest SEP level $\tilde{a} < \tilde{a}_0$ if $E[s_i|\tilde{a}_0] < \tilde{a}_0$.

The idea behind Definition 2 is simple. If the average memory induced by self-esteem $\tilde{a}_0$ is higher (respectively lower) than $\tilde{a}_0$, then the agent's self-esteem will drift up (respectively down), so that his self-esteem converges to the lowest SEP above (respectively the highest SEP below) $\tilde{a}_0$. Suppose, for instance, that in Example 1 the agent starts off with seed self-view $\tilde{a}_0 = 0.4$. The average experience he recalls is then below 0.4, lowering his self-esteem. This lowers the average experience he recalls, lowering his self-esteem further. Through this process, he converges to the low SEP.[3]

This definition allows us to make predictions about how temporary shocks affect the person's self-esteem, and to extend the static model to a dynamic one by assuming that the seed self-view in a period is determined by the previous period's SEP and a shock. Given this perspective, whenever there are multiple SEPs there is scope for the agent's self-esteem to shift drastically without any new information. For instance, if in Example 1 the agent starts off in the high SEP, but a temporary shock lowers his seed self-view to 0.4, then his self-esteem collapses to the low SEP. This motivates our final definition:

Definition 3 (Fragility)**.** A person has fragile self-esteem if there are multiple SEPs.

We assume that the agent understands that there might be multiple SEPs; for example, if he is in a low fragile SEP, he realizes that there is also a high SEP. Crucially, we also assume that this realization does not change his self-esteem. We believe this accurately reflects the underlying

[3] For formal convenience, the definition imposes that if the agent's seed self-view is exactly at an unstable solution to $\tilde{a} = E[s_i|\tilde{a}]$, then he ends up at the SEP above it.

psychology: when in a 'down' mood, thinking gloomy thoughts, it is impossible to see oneself in a positive light, even if one recognizes that at other times one has been, and might be again, in a situation characterized by positive thoughts.

For simplicity, we assume in the rest of the paper that there are at most two SEPs — a high and a low SEP. We identify conditions on the primitives for this to be the case:

Proposition 1. *Suppose that* $g(x) \leq g''(x)$ *and* $g'(x) \leq g'''(x)$ *for all* $x > 0$*. Then, there exist at most two SEPs.*

Roughly speaking, Proposition 1 holds when the peak of $g(x)$ is sharp enough: $g(x)$ is sufficiently convex (relative to its value $g(x) > 0$) and decreases quickly (relative to its third derivative). For example, $g(x) = e^{-x}$ satisfies the conditions in Proposition 1. Note also that this is only a sufficient condition; Example 1 does not satisfy it, but still has the property that there at most two SEPs.

In our model, the central feedback process arises from the effect of self-esteem on memory (as well as the more obvious effect of memories on self-esteem). There are, however, other feedback mechanisms that could plausibly lead to multiple equilibria. Some of these mechanisms are, like memory, what could be called "internal." For example, anxiety about the potential for poor performance could give rise to poor performance in domains such as public speaking (e.g., dry mouth), athletics, acting (e.g., forgetting one's lines), sex (anxiety-induced dysfunction) or exam-taking (Ariely et al., 2009), creating a direct link between confidence and performance similar to that in Compte and Postlewaite (2004). It is a perverse feature of human existence that so many mechanisms can come into play to undermine us at the very moment it is most important to perform optimally.

Other mechanisms involve external agents. For example, an individual who feels confident may project confidence to others, which can cause them to respond in a trusting, admiring, fashion, thus reinforcing the individual's confidence (Lamba and Nityananda, 2014, Johnson et al., 2011, Johnson and Fowler, 2011). Of course, by the same token, a lack of projected confidence can also usher forth a self-reinforcing response from others. None of these other mechanisms precludes the simultaneous operation of the memory-based mechanism we have discussed in detail. Indeed, these alternative mechanisms may reinforce the memory-based one.

Our model could also be extended to include appraisals of other aspects of life than self-esteem. Virtually the same framework could be applied to other self-relevant evaluations such as "How good is my life" (which might correspond closely to "life satisfaction") or "How good is the world," to the degree that they produce self-reinforcing feedback effects. By the same token, the model could be applied more narrowly to a particular type of self-evaluation — e.g., "How moral am I?" In all of these cases, one can easily imagine that the individual has a bank of memories relevant to the evaluation in question, and that their self-evaluation affects sampling from that bank.

3 Basic Properties of Self-Esteem

In the next three sections, we identify formal implications of our model. We take the perspective that self-esteem is an important component of utility, and hence we view self-esteem determination as a relevant object to study in itself. Accordingly, we first study basic properties of self-esteem.

3.1 Fragility of Self-Esteem

We begin by identifying the effect of the importance of self-esteem ($k > 0$) on the fragility of self-esteem. For many of our results on large k, we make the following assumption on $g(\cdot)$:

Assumption 1. $\lim_{x\to+\infty} g(x)x = 0$.

Assumption 1 implies that memories very far from one's current mood have a vanishing effect on self-esteem; that is, the memory process is quite sensitive to mood. Note that Example 1 satisfies Assumption 1.

Proposition 2 (Importance and Fragility of Self-Esteem)**.**

I. If $a = 0$ or $a = 1$, then the agent does not have fragile self-esteem.

II. For any $a \in [0, 1]$, if k is sufficiently small, then the agent does not have fragile self-esteem.

III. Suppose Assumption 1 holds. For any $a \in (0, 1)$, if k is sufficiently large, then the agent has fragile self-esteem.

Part I says that if the agent's evidentiary base is completely homogenous, then he cannot have fragile self-esteem. In this case, he always recalls the same experiences, so there is no scope for mood to

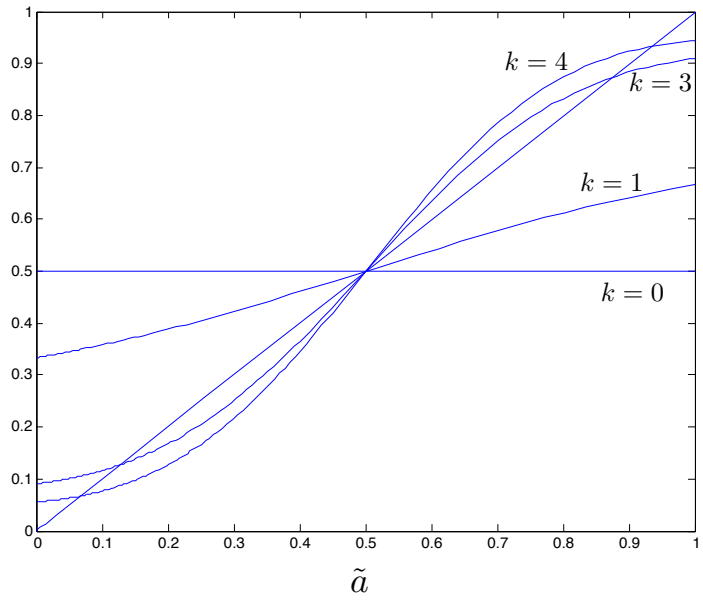

Figure 2: Importance k and Fragility of SEP

influence his memory. Part II says that if the agent's perceived ability is not that important, then he cannot have fragile self-esteem. In this case, his mood cannot much affect the type of experience he recalls. Conversely, Part III says that if the agent considers ability to be sufficiently important, then he is guaranteed to have fragile self-esteem. In this case, as a change in his self-esteem alters what facts he recalls, his mood and therefore his memories react strongly, making him sensitive to the self-reinforcing mechanism that generates fragility.

Figure 2 illustrates Parts II and III by modifying the example shown in Figure 1 with different levels of k. For $k = 0$, $E[s_i|\tilde{a}]$ is simply a horizontal line located at $a = 1/2$, the realistic level of self-esteem. For $k = 1$, the curve is no longer flat, but there is still only one SEP. For higher k, however, multiple SEPs arise.

3.2 Realism of Self-Esteem

Research in psychology has identified systematic biases in individuals' assessments of their own abilities, going in both directions. Research on overconfidence has focused on situations in which people

overestimate their own abilities (Camerer and Lovallo, 1999, Barber and Odean, 2001, Scheinkman and Xiong, 2003, Malmendier and Tate, 2005). But, a smaller body of research on the "impostor syndrome" has documented the opposite pattern (Langford and Clance, 1993, Sakulku, 2011, Young, 2011). Furthermore, researchers have documented these patterns for aspects of ability — such as intelligence or work performance — that are important to people. From a classical perspective, one might expect individuals to be best calibrated for the things that are most important to them. Our model helps to show why both types of systematic deviations might arise, and has the important implication that such deviations are most likely to occur when the individuals care a lot about their abilities.

Proposition 3 (Realism of Self-Esteem)**.**

I. For any $a \in (0,1)$, if k is sufficiently small, then the agent's SEP is realistic for a symmetric evidentiary base ($a = 1/2 \Rightarrow \tilde{a} = a$), unrealistically high for a left-skewed evidentiary base ($a > 1/2 \Rightarrow \tilde{a} > a$), and unrealistically low for a right-skewed evidentiary base ($a < 1/2 \Rightarrow \tilde{a} < a$).

II. Suppose Assumption 1 holds. For any $a \in (0,1)$, if k is sufficiently large, then there is a SEP with unrealistically high self-assessment ($\tilde{a} > a$) and a SEP with unrealistically low self-assessment ($\tilde{a} < a$), but there is no SEP with realistic self-assessment ($\tilde{a} = a$).

III. The set of SEP approaches $\{a\}$ as $k \to 0$. Also, if Assumption 1 holds and $a \in (0,1)$, then the set of SEP approaches $\{0,1\}$ as $k \to \infty$.

Part I of Proposition 3 can be interpreted as identifying the effect of skew in the evidentiary base on a person's self-assessment when k is small and hence fragility is not part of the picture. In particular, the agent underweights relatively rare observations. With a symmetric evidentiary base ($a = 1/2$), a realistic self-assessment does not bias the agent's memory in either direction, so the memories he recalls are consistent with his self-assessment. For a better evidentiary base ($a > 1/2$, i.e., left skew), a unique SEP must have $\tilde{a} > 1/2$, as for $\tilde{a} = 1/2$ the memories are on average better than $1/2$. Hence, the agent is more likely to recall good than bad memories, leading to a positive bias in his self-assessment.

Part II of the proposition says that if ability is sufficiently important, then the agent cannot have a realistic self-assessment, but can have both unrealistically low and unrealistically high self-

assessments. Note that this is true even if the evidentiary base is symmetric ($a = 1/2$). When the agent is self-assured, then (for k large) he is much more likely to recall good memories than bad memories, confirming his positive self-view. And when the agent is underconfident, then he is much more likely to recall bad memories than good memories, again confirming his self-view. Finally, Part III identifies limit versions of this point: when k is small, a person's mood does not affect the recall process much, so his self-assessment is close to the realistic one; but when k is large, his self-assessment becomes completely dissociated from evidence, becoming almost the highest possible or almost the lowest possible.

3.3 Vulnerability to Shocks

The ironic conclusion of the previous two subsections is that when self-esteem is important, it is also fragile and unrealistic. We now discuss an interesting relationship between these two phenomena: although self-esteem is unrealistic, its *degree of fragility* can reveal a lot about how unrealistic it is. To do so, we analyze the sensitivity of the agent's self-esteem to new experiences. A new experience could be an economic outcome or other event that the individual experiences at the moment and that changes how he feels — or, equivalently, what he recalls — about himself. When self-esteem is fragile, such seemingly small and/or temporary experiences can drastically affect a person's self-esteem and hence his well-being.

To illustrate, recall our Example 1, where the agent's realistic self-assessment is $a = 1/2$, but nevertheless both an extremely high and an extremely low self-esteem are SEPs. Suppose that the agent has converged to the high SEP in one period, and in the next period the way he thinks and feels about himself is negatively shocked to determine his new seed self-view. For a small negative shock, the decisionmaker returns to the high SEP. For a sufficiently large negative shock, however, his self-esteem collapses to the low SEP.

The point that people can slough off a certain amount of bad shocks, but when a threshold is exceeded they are likely to experience a large and abrupt plummeting of mood, has implications for the consequences of a sequence of shocks to self-esteem. If adjustments to shocks take time, then the impact of such a sequence can depend on whether the shocks occur one after the other in

quick succession or separated by a substantial time delay. An individual who experienced one shock each month, for example, might have sufficient time to return to the high SEP prior to receiving each subsequent shock, leaving him at the high SEP at the end of the sequence. In contrast, an individual at a high SEP who experienced a series of negative shocks in a short period of time would be less likely to have time to re-equilibrate between shocks, and, as a result, would be more likely to end up at the low SEP. This analysis predicts that an individual at a high SEP will be reluctant to experience a series of events with unpredictable consequences for mood in close succession, such as receiving a series of test results, since there is little upside, but a substantial downside if the shocks all turn out to be negative. Such an individual would be less averse to experiencing a sequence of shocks separated by time. Although we are not aware of direct tests of these predictions in the domain of self-esteem, closely related logic is at play in psychology accounts of stress.[4]

Beyond implying that a person can be sensitive to shocks, our model says that this sensitivity is quite informative. We define the *stability* of an SEP as the infimum of the shock sizes (in absolute value) that can lead the agent to shift to the other SEP. And we call an SEP less fragile, or more stable, if its stability is greater. As an illustration, Figure 3 varies a in Example 1. For $a = 1/2$, the curve $E[s_i|\tilde{a}]$ is symmetric, so the two SEPs are equally fragile. If $a = 0.4$, the two SEPs barely change, but their fragility greatly changes: now the low SEP is more stable and the high SEP is less stable. That is, the more justified the agent's high self-esteem is, the more stable it is to shocks; and conversely, the flimsier is the evidence on which a person's high but fragile self-esteem is based, the more fragile the self-esteem is. Intuitively, if the memories a low-ability but high-SEP agent recalls drop by a bit, the large share of bad experiences he has ignored come flooding back to his memory, setting him off on a tailspin. Generalizing:

Proposition 4. *Suppose there are two SEPs. An increase in a makes the high SEP more stable*

[4] The dominant account of stress in the psychological literature, for example, assumes that stress occurs when the challenges facing an individual exceed the protective capabilities of the 'coping' resources that the individual can marshal (Lazarus and Folkman, 1984). There is also some support for the idea that past stressors can cumulate and interfere with coping (e.g., Brewin et al., 2000, Turner et al., 1995). One important study (Bonanno et al., 2007) reported results from a phone study conducted in the New York City area after the September 11, 2001, terrorist attack. People who reported post-traumatic stress disorder symptoms (as compared with those who did not) differed on a number of dimensions, including demographic characteristics, close connection to the traumatic events, and, most important for this prediction, the occurrence of recent life stresses.

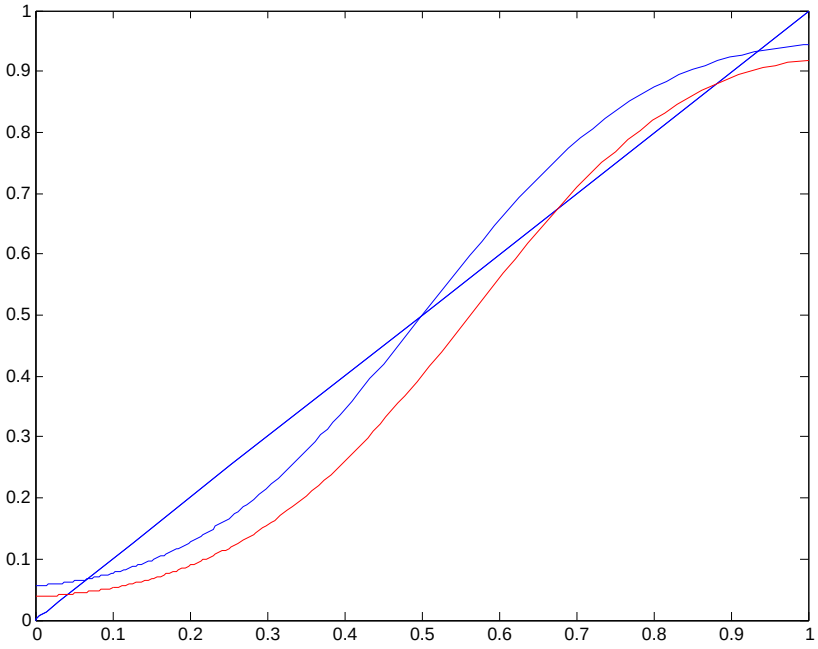

Figure 3: Realism and Fragility of SEP

and the low SEP less stable.

An interesting version of this point can obtain in the limit when k is large. Then, Part III of Proposition 3 implies that for any true ability a, self-esteem $\tilde{a}$ is close to either 0 or 1. Hence, the agent's self-esteem is not a reliable measure of his ability. But the fragility of the two SEPs can be informative about the agent's true ability. For instance, in Example 1 the stability of the high SEP approaches a as k approaches infinity, and the stability of the low SEP approaches $1-a$, so the fragility of self-esteem is perfectly informative about a.[5] Of course, in the current version of the model, it is unclear how an outside observer might observe fragility. But as we show below, the fragility has behavioral manifestations when the agent also takes actions to protect his self-esteem.

[5] To derive the unstable fixed point of $E[s_i|\tilde{a}]$ in Example 1, we solve out $\tilde{a} = E[s_i|\tilde{a}] = \frac{a}{a+\frac{g(k\tilde{a})}{g(k-k\tilde{a})}(1-a)} = \frac{a}{a+\frac{1+k^2(1-\tilde{a})^2}{1+k^2\tilde{a}^2}(1-a)}$. For a sufficiently large k, this is approximated by $\tilde{a} = \frac{a}{a+\frac{(1-\tilde{a})^2}{\tilde{a}^2}(1-a)}$, which can be rewritten as $a\tilde{a} + \frac{(1-\tilde{a})^2}{\tilde{a}}(1-a) = a$ or $(1+\frac{1-a}{a})\tilde{a}^2 - (1+2\frac{1-a}{a})\tilde{a} + \frac{1-a}{a} = 0$. By solving this, we obtain that the unstable fixed point is $1-a$.

3.4 Implications for Hedonics

Before we turn our attention to how self-esteem affects behavior, we briefly discuss the implications of the above results for a person's self-esteem after experiencing an outcome. Following literatures in economics as well as psychology, we take the view that self-esteem is a major component of well-being, and hence these effects are important in themselves.

The effects of a new experience can be motivated by a point Thomas Schelling (1987) makes in his brilliant essay "The Mind as a Consuming Organ." He describes how his mood, after giving a public lecture, depends critically on his appraisal of the quality of his performance, and he recognizes how irrational this is:[6]

> At my age the statistical record of my performance ought to reflect so many observations of good, poor, and mediocre performance that one more experience at either tail of the distribution could hardly affect a rational self-assessment. I try to remind myself that on those occasions when feedback depresses me; but my welfare function is apparently not constructed that way. It feels to me as if I am taking the audience reaction as evidence, and what makes me feel good or bad is the belief that my average career performance is high or low.

While Schelling's emotional reaction makes no sense when viewed from a standard perspective, the feedback processes embedded in our model provide a natural account of it. As shown in Figure 4, in our model experiences influence utility in three ways. The first is by directly influencing mood (path a). While in a classical model Schelling's thousandth lecture could plausibly influence his utility, in such a model, this would be the end of the story. Furthermore, any such direct effect on utility is either small or short-lived. Specifically, if instantaneous utility is determined by a combination of many experiences, then the effect a single talk is likely to be thinned out by other recent experiences and should be small; and if instantaneous utility is dominated by only one or

[6] A similar observation was made much earlier by William James, the path-breaking American psychologist, who actually coined the term "self-esteem." He wrote that "we ourselves know how the barometer of our self-esteem and confidence rises and falls from one day to another through causes that seem to be visceral and organic rather than rational, and which certainly answer to no corresponding variations in the esteem in which we are held by our friends" (James, 1890, page 307).

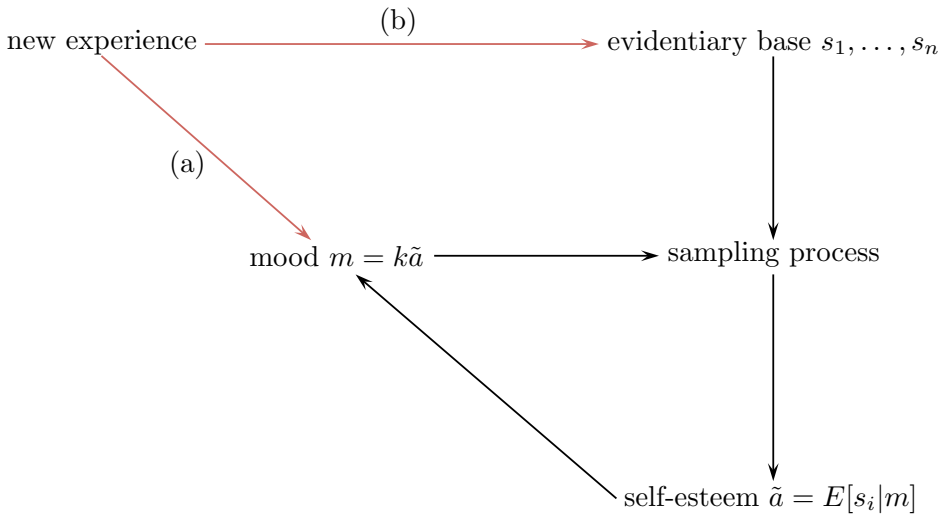

Figure 4: The Effects of a New Experience

two of the most recent experiences, then the effect of a talk is quickly overtaken by other, newer experiences and is therefore short-lived.

In our model, however, a new lecture can have other effects. The second is an indirect effect that acts through the influence of mood on the sampling process from the evidentiary base. Schelling's immediate emotional reaction to the recent lecture can constitute a new seed self-view which, if it leads to a shift past the unstable equilibrium point, can have a dramatic impact on self-esteem. As a result of this path, for an individual who is at a high but fragile SEP, the effects of a negative experience can be catastrophic in the sense of causing a decline that is totally out of proportion to the duration-weighted severity of the episode itself. (The reverse is, of course, true of a positive experience for an individual at a low but fragile SEP.)

Third, and important for people who unlike Schelling, have a limited bank of relevant experiences, a new experience can affect utility by becoming an element in the evidentiary base (path b). Even when a new experience does not immediately dislodge the individual from an existing SEP, such a new experience can still change the fragility of that equilibrium.

Although somewhat outside of the scope of the model, which focuses exclusively on ego-relevant

thoughts and behaviors, the combination of these effects helps to explain a phenomenon of momentous importance in the real world that standard economic models fail to account for: the potentially catastrophic impact of a traumatic event. Traditional models, in which overall utility is the integral of momentary utility experienced over time, cannot make sense of the enormously destructive impact such events can have, even when they are extremely short-lived. Our model, in contrast, provides a more compelling account of such an impact. As outlined in Figure 4 and discussed above, a traumatic experience can have the immediate effect of driving an individual who was at a high SEP to a low SEP. This effect is, however, likely to be transient. More importantly, a single traumatic even can enter into the evidentiary base. To analyze the impact in our model, we think of a trauma as an event that is much worse than the person's experiences so far. In our binary framework, we can capture this situation by assuming that a single 0 is added to an evidentiary base composed entirely of 1's. In combination, Parts I and III of Proposition 2 imply that if the event is particularly traumatic (i.e., large k), then it turns the agent's self-esteem from non-fragile to fragile. By implication, a single trauma renders the agent vulnerable to collapses in the evaluation of himself — or, more generally interpreted, his life. When such collapses occur, memories of the traumatic event dominate his thinking, in effect damning him to relive the experience. Such an account is consistent with a wide range of research documenting selective attention and memory in victims of PTSD (see, e.g., Buckley et al., 2000). Thus, the first 4 items in the "PTSD checklist" (Blanchard et al., 1996) all deal with selective recall of memories: intrusive recollections, flashbacks, being upset by reminders, and experiencing distressing dreams.

4 Self-Esteem-Dependent Choices

So far, we have considered how the agent's self-esteem is determined. The ultimate goal of the paper, however, is to analyze the implications of our framework for economically important decisions. By standard economic logic, a person's self-view affects decisions he makes. A researcher's belief about his ability, for instance, affects his assessment of the chances of success in different research projects, and hence it affects the difficulty or ambitiousness of the projects he chooses to engage in. At the same time, we take the perspective that the agent cares about his self-esteem, and hence

is motivated to take actions that increase his self-esteem. A researcher with high self-esteem, for instance, may be aware that his self-esteem can collapse, so he might avoid situations in which his ability is questioned. While in general the causality can run in both directions simultaneously, from self-esteem to the choices people make, and from choices to self-esteem, for simplicity of exposition we separate these two directions. We now discuss how the agent's self-esteem affects the choices he makes, assume that the action itself does not affect self-esteem. Section 5 analyzes models in which people take actions to increase self-esteem.

To analyze the consequences of insecurity of decisions involving work effort, we assume that the agent's intrinsic utility is $kae - c(e)$, where a and k are his ability and importance of the ability as previously, $e \geq 0$ is his effort, and $c(\cdot)$ is a strictly convex cost of effort with $c(0) = c'(0) = 0$. We think of kae as the agent's achievement or expected achievement given his true ability and effort. The agent's perceived ability $\tilde{a}$ is determined in an SEP, with k representing the importance of ability as defined in Section 2. This is consistent with the interpretation of k as the importance of ability in future life outcomes. The agent chooses his level of effort to maximize his perceived utility $k\tilde{a}e - c(e)$. Hence, the agent's effort is strictly increasing in his perceived ability, and it is optimal if and only if his perceived ability equals his true ability.

Two remarks regarding our model are in order. First, a key assumption is that higher self-esteem raises effort. In our framework, this arises because ability and effort are complements in production. Implicit in this interpretation is an assumption mentioned above: that — even though he understands the possibility of multiple SEPs — the agent views the self-assessment he would have in another SEP as unrealistic, and hence chooses his effort based on his current self-assessment. An alternative account is that the individual's energy level and motivation are directly positively affected by self-esteem or mood. This requires no assumption on the complementarity between ability and effort (or whether the agent views his current self-esteem as realistic), and has the same key implication that effort increases in self-esteem.[7] Second, note that we have not included utility from self-esteem in the agent's utility function. In the current model, this would make no

[7] Formally, suppose that intrinsic utility is $k(a+e) - c(\tilde{a}, e)$. The predictions of this formulation differ in only two ways from those below. First, in this alternative effort is always optimal, whereas in the above version it is often too high or too low. Second, in this version the agent's welfare is always higher in the high than in the low SEP, whereas in the above version it might not if his action is more distorted in the high than in the low SEP.

difference, as the agent cannot affect his self-esteem. Utility from self-esteem will, however, be a central consideration in Section 5.

An immediate implication of Proposition 3, which states that the agent cannot have a realistic self-assessment if k is sufficiently high, is that the effort she exerts will either be too low or to high, but never at the optimal level. To go further, we analyze the effect of incentives on effort choice. In a classical model, an increase in incentives always leads to an increase in effort. This is true in our model only in the range where stakes are low:

Proposition 5.

I. For any $a \in (0,1]$, if k is sufficiently small, then the agent's effort is strictly increasing in k.

II. Suppose Assumption 1 holds. If k is sufficiently large and the agent is in the low SEP, then his effort can be strictly decreasing in k.

III. Suppose Assumption 1 holds and $g(x)x$ is decreasing for sufficiently large x. For any $a \in (0,1)$, if k is sufficiently large and the agent is in the high SEP, then the agent's effort is strictly increasing in k.

As we have seen in Proposition 2, when k is low, the agent's SEP is unique. Though his SEP can be either increasing or decreasing in k in this case, $k\tilde{a}$ is strictly increasing in k. Intuitively, if k is sufficiently small, then the agent's recall process is not much affected by k, so an increase of k does not change his self-esteem much. Then, the direct effect of increasing k dominates, and the agent reacts in a classical way to incentives.

Proposition 2 also implies, however, that if k is sufficiently large, then the agent's self-esteem is fragile. In this range, an increase in incentive can *lower* the effort of an agent in the low SEP. To see this, recall our Example 1 with $a = 1/2$ and different levels of k. If $k \leq 2$, there is a unique SEP in which $\tilde{a} = \frac{1}{2}$, and an increase in incentives always increases the agent's effort. If $k > 2$, however, the solution $\tilde{a} = \frac{1}{2}k$ to the equation $\tilde{a} = E[s_i|\tilde{a}]$ becomes unstable, and there are two SEPs in which $\tilde{a}^{low} = \frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{k^2}}\right)$ and $\tilde{a}^{high} = \frac{1}{2}\left(1 + \sqrt{1 - \frac{4}{k^2}}\right)$. Also, $\tilde{a}^{low}$ is strictly decreasing in k and goes to zero as $k \to +\infty$, as shown Proposition 3 Part III. More interestingly, as we show in the proof of Proposition 5, $k\tilde{a}$ — and hence his expected achievement $k\tilde{a}e$ — is strictly decreasing in k in the range $k > 2$. Intuitively, as stakes increase an agent in the low SEP is less able to recall

positive facts about himself — these being very far — so his self-esteem decreases. This effect can outweigh the classical effort-increasing effect of the increase in stakes, reversing the classical prediction that the agent should exert more effort when incentives increase.

When applied to a single individual over time or across situations, the above logic says that low-self-esteem individuals may be most pessimistic about themselves, and as a result perform the worst, for important tasks or tests. In this sense, our model predicts a kind of choking, although through a mechanism that is different from those proposed in the psychology literature (Baumeister 1984 and Baumeister and Showers 1986; and in economics see Ariely et al. 2009).

In contrast, Part III says that, with the additional condition on $g(x)x$, the possibility of choking applies only to agents in a low SEP, so that our framework predicts that only low-self-esteem individuals can react in a perverse way to incentives. As stakes increase, an agent in a high SEP is less likely to recall bad facts about himself, increasing his self-confidence. This reinforces the classical effect of incentives.

5 Self-Esteem-Influencing Choices

We now analyze actions that are motivated by a desire to increase self-esteem. In our model, there are two primary ways in which a person might go about improving self-esteem. First, he can take actions to improve his mood, thereby potentially influencing which SEP she ends up in. Second, he can take actions to improve the evidentiary base by adding favorable new observations to it. Note, however, that for many or most of the important aspects of self-esteem, most people have an extensive evidentiary base. If this is the case, then adding further observations to it is likely to have a small effect in itself. Hence, we focus on the first channel. We model situations in which a shock occurs to the agent, and he realizes that this shock will have an effect on his seed self-view. He can, however, take preemptive actions that mitigate the effect of the shock. For example, consider an academic who will soon give a seminar on a new paper. He knows that some critical and outspoken audience members are likely to object to any shortcomings in a harsh way, and set him off recalling similar previous experiences. To prevent this, the researcher can take extra effort to prepare for the seminar, anticipating and forestalling any objections.

As the example of the academic illustrates, an important issue is how people feel about changes in outcomes caused by changes in effort. If accomplishment is a function of both ability and effort, then the key feature of the context will be whether it is accomplishment per se that is valued, or rather ability (or, possibly, effort). At one extreme, where ability is all that matters, any accomplishment that can be attributed to effort will be discounted. Some universities have the reputation of valuing only brilliance, in which case success that is clearly due to effort does not make researchers feel good about themselves. At other universities, however, there may be even a positive value to effort. It may also be the case that self-esteem depends only on ability, but after exerting a lot of effort, the person — naively or self-servingly — behaves as if a good outcome is due to ability. By incorporating a psychological factor which represents how much a person discounts own effort, denoted by $\delta \in [0, 1]$, we can easily accommodate these different possibilities about how effort affects feelings about oneself through outcomes.

Formally, we suppose for notational simplicity that $k = 1$. The agent starts off at a mood $\tilde{a}_{-1}$. He receives a shock to his mood, $\epsilon \in [-1, 1]$, where we posit that ϵ has full support to make sure that the agent can start in a full range of new moods. After observing the shock, the agent can take an action $e \geq 0$ with the effort cost $c(\cdot)$ that is strictly convex and $c(0) = c'(0) = 0$. The shock and effort determine his seed self-view according to $\tilde{a}_0 = \tilde{a}_{-1} + \epsilon + \delta e$. Returning to the example of the researcher, ϵ is his perception of the reactions he will face, and e is the effort he makes to give a more convincing seminar. The outcome of the seminar is then $\epsilon + e$. He might, however, discount the preparation effort, feeling that the seminar was convincing not because the research is promising, but because he overprepared. Hence, his mood is shocked by $\epsilon + \delta e$ to give his seed self-view. We could assume that the effect of effort is probabilistic; this would not change the qualitative conclusions regarding the agent's willingness to exert effort, but it would mean (realistically) that the agent cannot guarantee himself a high SEP. Finally, the agent's SEP is determined as in our basic model with seed self-view $\tilde{a}_0$, and since our primary goal is to analyze actions motivated by the desire to increase self-esteem, we assume that the agent's goal is to maximize his resulting level of SEP self-esteem minus the cost of effort.

Proposition 6. *I. If the agent is not fragile or $\delta = 0$, then $e = 0$.*

II. If he is fragile and $\delta > 0$, then (i) the average action is positive; (ii) given $\tilde{a}_{-1}$ and ϵ, an increase in a either discontinuously increases e from zero or continuously strictly decreases e.

Given that the only motive to exert effort is to influence his self-esteem, if the SEP is unique or he fully discounts effort, then the agent does not exert effort. More interestingly, if the agent's self-esteem is fragile and he does not fully discount effort, then he exerts positive effort in expectation. Furthermore, if the effort he exerts is positive, then it is strictly decreasing in his true ability. This formalizes the intuition that being desperate to increase one's self-esteem is a sign of low ability. Intuitively, an increase in ability makes the low SEP less stable and the high SEP more stable. As a result, it is easier to manipulate oneself to a high SEP. Similarly, if the agent's starting self-view is less favorable, then it takes more to achieve a high SEP, so the agent either ceases to exert effort or exerts greater effort to achieve the high self-esteem. Proposition 6 potentially captures a wide range of behaviors people perform to increase self-esteem. While it is difficult to conclusively rule out motives for specific actions based on classical instrumental utility, individuals often appear to exert effort that is counterproductive or out of proportion from the perspective of classical utility. For instance, when a rich person works ceaselessly at the expense of family life, he is probably doing so to feel good about himself rather than to improve his material payoffs.

Of course, in the current model the agent maximizes utility taking self-esteem into account. But more generally interpreted, he might be maximizing short-term utility, and for well-known reasons, this may involve underinvestment in long-term objectives (e.g., Laibson, 1997, O'Donoghue and Rabin, 1999). Even if a workaholic feels better in the short run from neglecting his family at the expense of work performance, this might be suboptimal from the perspective of discounted utility.

From the perspective of the current model, the agent has an incentive to increase his seed self-view either in a low SEP or in a high SEP: if in a high SEP, he is willing to work hard to try to maintain a high SEP, and if in a low SEP, he is willing to work hard to try to shift to a high SEP. For instance, a workaholic might work hard both to retain and to obtain a high SEP. If either effort and ability are complements (as in Section 4) or the cost of effort is decreasing in how he feels about himself, the latter type of behavior is less common than the former type.

A direct effort to increase mood is not the only way a person can seek high self-esteem. Suppose

that seed self-view is given by $\tilde{a}_{-1} + (1+e)\epsilon$, where $e \in [-1, 1]$ is the effort level the agent exerts to decrease or increase his exposure to the shock ϵ (e.g., by actively avoiding or seeking out a respected colleague's feedback regarding his presentation). Our model predicts that if the agent is in a high but fragile SEP, then he is willing to incur costs to reduce his exposure to shocks that could lower his memories. Conversely, if he is in a low but fragile SEP, then he is willing to incur costs to increase his exposure to shocks. Once again, under the assumption that the absolute value of effort is increasing in self-esteem, the former type of behavior is likely be more common than the latter.

The above insight leads to a new interpretation of a type of behavior referred to by psychologists as "self-handicapping." Self-handicapping arises when an individual faces a task, performance on which, they fear, could provide negative information. In one set of psychology experiments (Berglas and Jones, 1978), this situation was created experimentally by informing subjects that they had done well on a task that they believed they had done poorly on, making them feel that their high performance was a fluke that would be unlikely to be repeated. They were then given a choice of drugs to take ranging from those that were performance-enhancing to those that were performance-debilitating. Individuals put into a state of doubt about their own ability to repeat a prior favorable performance were more likely to take the performance-debilitating drug, thus virtually ensuring that the feared outcome would happen, but rendering it undiagnostic of ability. In other self-handicapping research, individuals whose confidence are shaken are asked to perform a task and allowed to determine the difficulty of the task. Self-handicapping involves choosing a task that is inappropriately easy, so the individual is virtually certain to perform well, providing little opportunity for learning about ability, or, conversely a task that is inappropriately difficult so that the individual can attribute their almost inevitable failure to the difficulty of the task (e.g., Greenberg, 1985). In all of these cases, to avoid self-relevant shocks, individuals engage in costly behaviors that lower their anticipated performance and, as a result, expected material payoffs.

Previous economic accounts can conceptualize self-handicapping as a form of information avoidance, whereby a person dislikes receiving a potentially useful signal because it depresses motivation (Bénabou and Tirole, 2002), lowers anticipatory utility or ego utility (Kőszegi, 2006), or because it leads to sensations of loss (Kőszegi and Rabin, 2009). While these motives are surely important,

they are calibrationally unrealistic accounts of a non-trivial preference for self-handicapping. Most people have plenty of information about themselves, so a minor piece of news is likely to affect the sum-total of evidence only by a trivial amount. Hence, these accounts should not predict a strong motivation to self-handicap. To the extent that a bad talk, a bad opinion by a friend, or another bad signal can affect the agent's memories, and perturb them from a high to a low SEP, however, a seemingly outsized motive to avoid the signal makes perfect sense in our model.

6 Applications

In this section we outline a number of applications of our framework, to education, job search, workaholism and aggression. For the first two of these, we argue that suboptimal effort can occur in two, quite different situations: people who are in a low SEP and who underestimate their own ability to succeed and hence their expected return to effort, and people at a high SEP who are reluctant to engage in actions that could put their high SEP at risk. Because all situations we consider pertain to important areas of life, we focus on implications of our results when k is high so that self-esteem is fragile.

6.1 Education

The rate of return on education is high for many who choose not to continue their schooling or who put minimal effort into studying. Under conventional economic analysis — under which investment in education depends on an individual's expected discounted returns to schooling and studying — the main individual-level predictors of under-investment should be high discount rates, low levels of self-confidence, and pessimistic beliefs about the returns to education.

Yet there is some evidence that under-investment is not due to either lack of information about the returns to education, to high discount rates, or to present bias. The informational account of under-investment would predict that giving people better information about the returns to education would increase investments in schooling, but tests of such information interventions have been mixed at best. Some studies, almost exclusively in developing countries, observe beneficial effects of informational interventions on both beliefs and outcomes (Jensen, 2010, Avitabile and

De Hoyos, 2018), some find effects only on beliefs (e.g., McGuigan et al., 2016), but others find little or no impact of informational interventions on either beliefs or behavior (e.g., Fryer Jr, 2013). Moreover, a common finding in this literature is that positive effects, to the extent that they are observed, tend to be limited to more able and affluent students.

Likewise, the view that under-investment arises from high discount rates or present bias has led to a variety of interventions that involve providing immediate incentives to compensate for the lack of motivational impact of distant, and hence discounted, benefits. Such interventions have yielded almost uniformly disappointing results (Fryer Jr., 2011, Leuven et al., 2010, Bettinger, 2012, Lavecchia et al., 2016), which calls into question the importance of time discounting as an explanation for under-investment.

With lack of information and steep time discounting explaining little of the under-investment in education, conventional logic suggests that lack of self-confidence must play a major role. Indeed, the first implication of our model — that people at low SEPs will be underconfident — is consistent with this account. Low investment might well make sense for those low in ability, but our analysis suggests that people who are at low SEPs will typically be under-confident — i.e., will underestimate their own abilities, and hence their own returns to education. Unlike explanations for low investment based on high discount rates, which could be argued to reflect true preferences for earlier gratification, or on low levels of self-confidence, which, in the standard model are assumed to reflect accurate assessments of true ability, the low levels of investment at low SEPs predicted by our model result from underestimate of ability, and hence can be interpreted as true cases of under-investment.

For students at high SEPs, the situation is different. Section 3 implies that students who are at high SEPs revolving around academic achievement will generally have inflated views of their own academic prowess, and by Section 4, they will tend to work excessively hard. This tendency might be exacerbated by their fear of losing their high SEP (Section 5): just like the workaholics below, they might exert supra-optimally high levels of effort in an attempt to avoid failure and the potential drop to a low SEP. At the same time, and also in an attempt to avoid a collapse in self-esteem, however, they might engage in tasks that are insufficiently ambitious and therefore do

not expose them to risk of failure.

Our framework also provides a new perspective on one of the most prominent set of findings in recent research on education: the effect of a 'fixed mindset,' which sees intellectual ability as innate and immutable, as compared with a 'growth mindset,' which sees intelligence as a capability that can be developed. Virtually all of the research finds that those with growth mindsets fare much better academically than those with fixed mindsets (e.g., Uluduz and Gunbayi, 2018), and the benefits of a growth mindset seem to be especially pronounced when students are confronted with setbacks such as a poor grade on a test. Those with fixed mindsets tend to take a failure to heart and assume it means they are not up to the task, whereas those with growth mindsets tend to respond by redoubling their efforts.[8]

Conventional economics provides a natural interpretation of this phenomenon: students who believe that there are high returns to effort will naturally put in greater effort. Our perspective also encompasses this effect, but identifies another mechanism through which mindset can matter. To define mindset in our model, we let a and e denote innate ability and effort, respectively, and assume that academic achievement depends on $ka + (1-k)e$. In this formulation, a fixed mindset corresponds to a high k, and a growth mindset to a low k. Then, students who have fixed mindsets may respond more negatively to a negative outcome such as an unfavorable test result because they will see it as reflecting on their core abilities. Such a negative emotional shock in our model, can then put them into a low SEP in which they will underestimate their own ability, especially if they have had no or few major failures, and therefore the new experience creates or reinforces a low SEP. For those with a growth mindset, in contrast, adverse information about performance will have a smaller negative emotional impact.

[8] In one of the few studies examining these issues conducted by economists, Alan et al. (forthcoming) randomly assigned some students in two elementary schools in Turkey to receive an intervention designed to induce "grit." As described by the researchers, the intervention "aims to positively influence children's beliefs about the malleability of ability and the productivity of effort in the skill accumulation process, and thereby induce gritty behavior. The program exposes children to a world view in which ability, rather than being innately fixed, can be developed through sustained, goal-oriented effort. The core message is to highlight the role of effort in the skill accumulation process and thereby in achievement, and to discourage students from interpreting early setbacks and failures as evidence for a lack of innate ability." The program yielded a variety of substantial benefits, including raising test scores and decreasing the students' likelihood of giving up after failure.

6.2 Job Search

Job search in standard economics is a fairly cut-and-dried affair: the job-searcher balances the expected gains from additional search against its marginal costs. The reality for many job-seekers, however, is difficult to reconcile with such an account. In diverse surveys, unemployment has been found to be one of the most severe and reliable causes of unhappiness (e.g., Clark and Oswald, 1994, Helliwell and Huang, 2014, Winkelmann and Winkelmann, 1998). Given how miserable people find the state of being unemployed, one might expect to observe huge amounts of time and effort applied to alleviating the state. Yet, the small number of studies that have focused on time-use by the unemployed have found that unemployed individuals devote surprisingly little time to job search. For example, one international comparison (Krueger and Mueller, 2012) found that average search time is highest in the United States, at half an hour per day, and that Europeans search much less, with Finland at the bottom at 3 minutes per day. These means are anchored by the large fraction of unemployed who don't search at all. Again, the U.S. is highest among countries studied, at 20% searching, and Finland is the lowest, at 5% (see, also, Krueger and Mueller, 2010).

To account for this and other commonly observed features of job search that do not fit the predictions of standard models, behavioral economists have proposed modified models that introduce psychological enrichments, such as biased expectations (e.g., Spinnewijn, 2015, Altmann et al., 2018), present-biased preferences (e.g., DellaVigna and Paserman, 2005), and reference-dependent preferences (e.g., DellaVigna et al., 2017), but none of these incorporate the self-esteem-related causes and consequences of job search.

As with education, high and low SEPs should produce different patterns of job search behavior. A job searcher who is at a low SEP, which one can assume many are, is not motivated to search because he perceives the returns to search to be low; indeed, according to our model, they will underestimate their own worth, and hence, likely, their probability of finding a job. This perception could translate into low job search, setting a low reservation wage, or both.

For individuals at a high SEP, the implications of the model are different. On the one hand, one might expect such individuals to be exerting supra-optimal effort given that we predict they will be overly optimistic about their prospects of finding a job. However, several factors could

mitigate against such an outcome. First, as with education, there is some chance that job searchers might feel so confident about getting a good job that they do not bother to put in much effort — although that seems unlikely in this context. More interestingly, and a more novel prediction that comes out of our framework, is that job search might be avoided if it threatens the stability of a high SEP. Searching for a job but not finding one, especially if there is explicit rejection (e.g., being told that one is not qualified for a specific job, as contrasted with simply not hearing back from an inquiry) can result in a negative signal that not searching for a job and hence not finding one does not. In this situation, in our model, an individual might not engage in search to avoid threatening a high but fragile SEP. Our model also predicts that an individual might engage in specific types of search activities that are less productive on the margin if they are also less likely to result in explicit rejection. A job searcher might, for example, spend a greater than optimal amount of time sifting through want ads or sending out resumes, as compared with activities such as making phone calls or interviewing for jobs, that can result in explicit rejection and hence potentially perturb a high but fragile SEP.

Yet another prediction of our model that is not generated by conventional models is that, in line with our discussion after Proposition 4, an individual might be willing to search (and be rejected) a single, or only a limited number of times within a given period, but after reaching some threshold might choose to take a break — providing time to re-equilibrate to the high SEP — before searching further.

6.3 Workaholism

The term "workaholism" was coined by psychologist Wayne Oates in a 1968 essay in which he confessed to being subject to the syndrome he gave a name to. (Tellingly, Oates authored 57 books over the course of his career.) Oates identified workaholism as a disorder akin to substance abuse, and contemporary definitions of workaholism often include components that overlap with definitions of addiction, such as "working beyond what is reasonably expected of the worker (as established by the requirements of the job or basic economic needs) despite the potential for negative consequences (e.g., marital issues)" (Clark, 2016). Much, although not all, subsequent research on

workaholism has continued to use the term to refer to a negative syndrome that is distinct from "work engagement," a term commonly used to refer to a positive connection to work. Studies of workaholism have identified numerous negative consequences of the syndrome, including job stress, work-life conflict and burnout. A consistent finding from research on workaholism is that the long work hours are not associated with high individual productivity (Clark et al., 2016).

Why would a person work longer hours than they need to, at the cost of decrements to their well-being and that of their family, and even to their own productivity? Consistent with Oates's view, one possible explanation might come from economic models of addiction (e.g., Becker and Murphy, 1988, Orphanides and Zervos, 1995).[9] Our framework identifies an additional mechanism that could plausibly play a role. Individuals who are at high but fragile SEPs may exert effort to maintain their high SEP. Since fragility in our framework is exacerbated by low ability, our model would predict that workaholics would be, on average, of lower ability, which might, beyond the effect of inefficient work strategies, help to explain the low productivity consistently observed in the empirical research on the phenomenon.

Studies that have examined the personality correlates of workaholism have, not surprisingly, found that workaholism is positively related to achievement-oriented traits such as Type A personality, achievement motivation and perfectionism. Correlations with other commonly measured personality traits (e.g., the 'big five') have been much weaker and more variable across studies. Interestingly, however, and supportive of the fragile self-esteem account of the syndrome, several studies have found that narcissism, which is the trait that psychologists have most frequently connected to high but fragile self-esteem, is positively related to workaholism (Andreassen et al., 2012, Clark et al., 2010). Prior studies that have sought to measure the combination of high but fragile self-esteem with the goal of predicting aggression (see following subsection) have done so using scales designed to measure narcissism (e.g., Raskin and Terry, 1988, Alexander et al., 2010).

[9] In such models, the more one engages in the addictive activity, the lower are the returns to alternative activities. This description does, in fact, seem likely to apply to work. For example, developing friendships takes time, and time spent at work could very plausibly detract from friendships, family life, and other forms of social capital. This will naturally produce the self-reinforcing dynamic characteristic of many forms of addiction.

6.4 Aggression

Our fourth application, aggression, has received the lion's share of attention in the psychology literature dealing with insecurity. In a book titled *Evil: Inside Human Violence and Cruelty*, Baumeister (1997) argues that much if not most aggression results from people who "think well of themselves most of the time but who are vulnerable to frequent or large fluctuations in their self-esteem." For these people, Baumeister continues, "that miserable sinking feeling that goes with a drop in self-esteem is all too familiar, and they are on guard to avoid it. Even a slight hint or mild implication that questions their personal worth may elicit a strong response." Baumeister uses his theory to explain a wide range of stylized facts about violence, including why men are so much more violent than women and why a violent reaction is so much more likely when a perceived snub occurs in public. Translating Baumeister's insight into our perspective, an insecure individual will be especially prone to violence because they will be highly motivated to lash out against any entity that threatens to expose weaknesses — i.e., to add unfavorable information to, or highlight unfavorable data in, the individual's evidentiary base. The plethora of adages such as "They who Threaten are afraid" (Diary of Benjamin Franklin) and "All Cruelty Springs from Weakness" (Seneca) point to the ubiquity of the phenomenon. Of course, many if not most acts of aggression do not involve physical violence. In academia, for example, how many nasty referee reports reflect as much on the insecurities of the reviewer as on weaknesses in the paper under review?

Our model formalizes the central mechanism identified by Baumeister. Suppose that the agent's partner acts in a way that the agent perceives as potentially shining negative light on him. This ego threat is akin to a low shock ϵ in Section 5. The agent then aggresses — chooses a high e — which leads his partner to stop the behavior in question. This eliminates the negative light that was threatening the agent, improving his feelings about himself. Importantly, this requires the agent not to discount his aggression in evaluating what happened (i.e., that his δ is relatively high). If the agent understood that his partner changed her behavior merely because of his aggression, he could not have succeeded in improving how he feels about himself.

7 Concluding Remarks

In this paper we have sought to bring the tools of economics to bear on insecurity, a phenomenon to which psychologists have attributed a range of self-destructive and often other-destructive patterns of behavior. Psychological theories have generally been vague about exactly what it means to have high but fragile self-esteem — i.e., how and why such a state could occur, and how and why it leads to the consequences that are commonly attributed to it. Here, we attempt to formalize some of the findings in the psychology literature to develop a rigorous model of the causes and consequences of fragile self-esteem, and discuss its implications for important economic activities such as education and job search.

Although we have assumed throughout the paper that people experience overall levels of self-esteem that can be represented by a single scalar, it would be more realistic to recognize that self-esteem is typically multidimensional. The same individual might, for example, experience different levels of self-esteem for different aspects of life such as work, family, friends, athletics, and so on (Pelham, 1995), raising a number of interesting issues. An obvious implication is that people with more "diversified" self-esteem — one that is based on more dimensions of similar importance — will experience smaller swings of mood and overall self-esteem, and existing research in psychology provides empirical support for this prediction (Linville, 1985, 1987). Our model, however, provides some novel predictions and insights. First, our model predicts that such diversification not only reduces mood volatility through the usual mechanism, but will also, potentially, reduce the fragility of an individual's self-esteem in a fashion analogous to the accumulation of experience. Second, the mood-congruent memory mechanism at the core of our model helps to explain how and why, a change in self-esteem in one domain can spill over and affect self-esteem in others, which can reduce the benefits of diversification.

That there are multiple dimensions of self-esteem also raises the possibility that individuals can choose (consciously or subconsciously) the weight they attach to each dimension. Claude Steele, one of the foremost researchers on education-related topics, provides an important example in a thoughtful essay on "Race and the Schooling of Black Americans." He writes about a process he calls "disidentification," in which a black student, "at precisely the time when he would need to

see school as a viable source of self-esteem resists measuring himself against its values and goals. He languishes there [i.e., in school], held by the law, perhaps even by his parents, but not allowing achievement to affect his view of himself."

To focus on the implications of insecurity, per se, and to explore the range of implications we can derive from relaxing only the conventional assumption of stable self-esteem (or, more commonly, the complete neglect of the construct), we have not incorporated other, by now commonplace, behavioral elements into the model, such as present-bias, reference-dependence, and gain-loss asymmetries. Rigorous analysis of the interactions between any of them and mood-congruent memory would be fruitful. For example, if memories are sensitive not only to the level, but also to recent changes in mood, then the agent will be more susceptible to collapses in self-esteem than without reference dependence.

References

Alan, Sule, Teodora Boneva, and Seda Ertac, "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit," *Quarterly Journal of Economics*, forthcoming.

Alexander, G. Caleb, Jennifer Humensky, Cesar Guerrero, Hannah Park, and George Loewenstein, "Brief Report: Physician Narcissism, Ego Threats, and Confidence in the Face of Uncertainty," *Journal of Applied Social Psychology*, 2010, *40* (4), 947–955.

Altmann, Steffen, Armin Falk, Simon Jäger, and Florian Zimmermann, "Learning about Job Search: A Field Experiment With Job Seekers in Germany," *Journal of Public Economics*, 2018, *164*, 33–49.

Andreassen, Cecilie Schou, Holger Ursin, Hege R. Eriksen, and Ståle Pallesen, "The Relationship of Narcissism with Workaholism, Work Engagement, and Professional Position," *Social Behavior and Personality: An International Journal*, 2012, *40* (6), 881–890.

Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar, "Large Stakes and Big Mistakes," *The Review of Economic Studies*, 2009, *76* (2), 451–469.

Avitabile, Ciro and Rafael De Hoyos, "The Heterogeneous Effect of Information on Student Performance: Evidence From a Randomized Control Trial in Mexico," *Journal of Development Economics*, 2018, *135*, 318–348.

Barber, Brad M. and Terrance Odean, "Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment," *The Quarterly Journal of Economics*, 2001, *116* (1), 261–292.

Baumeister, Roy F., "Choking Under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance.," *Journal of Personality and Social Psychology*, 1984, *46* (3), 610.

_ , *Evil: Inside Human Violence and Cruelty*, W.H. Freeman, 1997.

_ **and Carolin J. Showers**, "A Review of Paradoxical Performance Effects: Choking Under Pressure in Sports and Mental Tests," *European Journal of Social Psychology*, 1986, *16* (4), 361–383.

Beck, Aaron T., *Cognitive Therapy of Depression*, Guilford press, 1979.

Becker, Gary S. and Kevin M. Murphy, "A Theory of Rational Addiction," *Journal of Political Economy*, 1988, *96* (4), 675–700.

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

Bénabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

Bénabou, Roland and Jean Tirole, "Mindful Economics: The Production, Consumption, and Value of Beliefs," *Journal of Economic Perspectives*, 2016, *30* (3), 141–64.

Berglas, Steven and Edward E. Jones, "Drug Choice as a Self-Handicapping Strategy in Response to Noncontingent Success.," *Journal of Personality and Social Psychology*, 1978, *36* (4), 405.

Bernheim, B. Douglas and Raphael Thomadsen, "Memory and Anticipation," *The Economic Journal*, 2005, *115* (503), 271–304.

Bettinger, Eric P., "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores," *Review of Economics and Statistics*, 2012, *94* (3), 686–698.

Blanchard, Edward B., Jacqueline Jones-Alexander, Todd C. Buckley, and Catherine A. Forneris, "Psychometric Properties of the PTSD Checklist (PCL)," *Behaviour Research and Therapy*, 1996, *34* (8), 669–673.

Blaney, Paul H., "Affect and Memory: A Review.," *Psychological Bulletin*, 1986, *99* (2), 229.

Bonanno, George A., Sandro Galea, Angela Bucciarelli, and David Vlahov, "What Predicts Psychological Resilience After Disaster? The Role of Demographics, Resources, and Life Stress.," *Journal of Consulting and Clinical Psychology*, 2007, *75* (5), 671.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Memory, Attention, and Choice," 2017. Working Paper.

Brennan, Geoffrey and Philip Pettit, *The Economy of Esteem: An Essay on Civil and Political Society*, Oxford University Press, USA, 2004.

Brewin, Chris R., Bernice Andrews, and John D. Valentine, "Meta-Analysis of Risk Factors for Posttraumatic Stress Disorder in Trauma-Exposed Adults.," *Journal of Consulting and Clinical Psychology*, 2000, *68* (5), 748.

Buckley, Todd C., Edward B. Blanchard, and W. Trammell Neill, "Information Processing and PTSD: A Review of the Empirical Literature," *Clinical Psychology Review*, 2000, *20* (8), 1041–1065.

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

Caplin, Andrew and Kfir Eliaz, "AIDS Policy and Psychology: A Mechanism-Design Approach," *RAND Journal of Economics*, 2003, pp. 631–646.

Carrillo, Juan D. and Thomas Mariotti, "Strategic Ignorance as a Self-Disciplining Device," *The Review of Economic Studies*, 2000, *67* (3), 529–544.

Clark, Andrew E. and Andrew J. Oswald, "Unhappiness and Unemployment," *The Economic Journal*, 1994, *104* (424), 648–659.

Clark, Malissa A., "Workaholism: Its Not Just Long Hours on the Job," *Psychological Science Agenda*, 2016.

_ , **Ariel M. Lelchook, and Marcie L. Taylor**, "Beyond the Big Five: How Narcissism, Perfectionism, and Dispositional Affect Relate to Workaholism," *Personality and Individual Differences*, 2010, *48* (7), 786–791.

_ , **Jesse S. Michel, Ludmila Zhdanova, Shuang Y. Pui, and Boris B. Baltes**, "All Work and No Play? A Meta-Analytic Examination of the Correlates and Outcomes of Workaholism," *Journal of Management*, 2016, *42* (7), 1836–1873.

Compte, Olivier and Andrew Postlewaite, "Confidence-Enhanced Performance," *American Economic Review*, December 2004, *94* (5), 1536–1557.

Dana, Jason, Roberto Weber, and Jason Kuang, "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness," *Economic Theory*, 2007, *33* (1), 67–80.

DellaVigna, Stefano and M. Daniele Paserman, "Job Search and Impatience," *Journal of Labor Economics*, 2005, *23* (3), 527–588.

_ , **Attila Lindner, Balázs Reizer, and Johannes F. Schmieder**, "Reference-Dependent Job Search: Evidence from Hungary," *The Quarterly Journal of Economics*, 2017, *132* (4), 1969–2018.

Farnham, Shelly D., Anthony G. Greenwald, and Mahzarin R. Banaji, "Implicit Self-Esteem," in D. Abrams M.A. Hogg, ed., *Social Identity and Social Cognition*, Blackwell Publishing, 1999, pp. 230–248.

Ferrari, Joseph R and Ted Thompson, "Impostor Fears: Links with Self-Presentational Concerns and Self-Handicapping Behaviours," *Personality and Individual Differences*, 2006, *40* (2), 341–352.

Golman, Russell, David Hagmann, and George Loewenstein, "Information Avoidance," *Journal of Economic Literature*, 2017, *55* (1), 96–135.

Greenberg, Jerald, "Unattainable Goal Choice as a Self-Handicapping Strategy 1," *Journal of Applied Social Psychology*, 1985, *15* (2), 140–152.

Grossman, Zachary and Joël J. van der Weele, "Self-Image and Willful Ignorance in Social Decisions," *Journal of the European Economic Association*, 2017, *15* (1), 173–217.

Helliwell, John F. and Haifang Huang, "New Measures of the Costs of Unemployment: Evidence from the Subjective Well-Being of 3.3 million Americans," *Economic Inquiry*, 2014, *52* (4), 1485–1502.

Hetts, John J., Michiko Sakuma, and Brett W. Pelham, "Two Roads to Positive Regard: Implicit and Explicit Self-Evaluation and Culture," *Journal of Experimental Social Psychology*, 1999, *35* (6), 512–559.

Higgins, Raymond L, Charles Richard Snyder, and Steven Berglas, *Self-Handicapping: The Paradox that isnt*, Springer Science & Business Media, 2013.

James, William, *The Principles of Psychology*, Vol. 1, Holt, 1890.

Jensen, Robert, "The (Perceived) Returns to Education and the Demand for Schooling," *The Quarterly Journal of Economics*, 2010, *125* (2), 515–548.

Johnson, Dominic D.P. and James H Fowler, "The Evolution of Overconfidence," *Nature*, 2011, *477* (7364), 317.

_ , Nils B. Weidmann, and Lars-Erik Cederman, "Fortune Favours the Bold: An Agent-Based Model Reveals Adaptive Advantages of Overconfidence in War," *PLOS One*, 2011, *6* (6), e20851.

Jr., Roland G. Fryer, "Financial Incentives and Student Achievement: Evidence from Randomized Trials," *The Quarterly Journal of Economics*, 2011, *126* (4), 1755–1798.

Jr, Roland G Fryer, "Information and Student Achievement: Evidence From a Cellular Phone Experiment," Technical Report, National Bureau of Economic Research 2013.

Kernis, Michael H., "Following the Trail from Narcissism to Fragile Self-Esteem," *Psychological Inquiry*, 2001, *12* (4), 223–225.

_ , **Teresa A. Abend, Brian M. Goldman, Ilan Shrira, Andrew N. Paradise, and Christian Hampton**, "Self-Serving Responses Arising from Discrepancies Between Explicit and Implicit Self-Esteem," *Self and Identity*, 2005, *4* (4), 311–330.

Kőszegi, Botond, "Ego Utility, Overconfidence, and Task Choice," *Journal of the European Economic Association*, 2006, *4* (4), 673–707.

_ , "Utility from Anticipation and Personal Equilibrium," *Economic Theory*, 2010, *44* (3), 415–444.

_ **and Matthew Rabin**, "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 2006, *121* (4), 1133–1166.

_ **and** _ , "Reference-Dependent Consumption Plans," *American Economic Review*, 2009, *99* (3), 909–936.

Koole, Sander L., Ap Dijksterhuis, and Ad Van Knippenberg, "What's in a Name: Implicit Self-Esteem and the Automatic Self.," *Journal of Personality and Social Psychology*, 2001, *80* (4), 669.

Kőszegi, Botond, "Health Anxiety and Patient Behavior," *Journal of Health Economics*, 2003, *22* (6), 1073–1084.

Krueger, Alan B. and Andreas I. Mueller, "The Lot of the Unemployed: A Time Use Perspective," *Journal of the European Economic Association*, 2012, *10* (4), 765–794.

_ **and Andreas Mueller**, "Job Search and Unemployment Insurance: New Evidence from Time Use Data," *Journal of Public Economics*, 2010, *94* (3-4), 298–307.

Laibson, David I., "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 1997, *112* (2), 443–477.

Lamba, Shakti and Vivek Nityananda, "Self-Deceived Individuals are Better at Deceiving Others," *PloS one*, 2014, *9* (8), e104562.

Langford, Joe and Pauline Rose Clance, "The Imposter Phenomenon: Recent Research Findings Regarding Dynamics, Personality and Family Patterns and their Implications for Treatment.," *Psychotherapy: Theory, Research, Practice, Training*, 1993, *30* (3), 495.

Lavecchia, Adam M., Heidi Liu, and Philip Oreopoulos, "Behavioral Economics of Education: Progress and Possibilities," in "Handbook of the Economics of Education," Vol. 5, Elsevier, 2016, pp. 1–74.

Lazarus, Richard S. and Susan Folkman, "Coping and Adaptation," *The Handbook of Behavioral Medicine*, 1984, *282325.*

Leuven, Edwin, Hessel Oosterbeek, and Bas Van der Klaauw, "The Effect of Financial Rewards on Students' Acheivements: Evidence From a Randomized Experiment," *Journal of the European Economic Association*, 2010, *8* (6), 1243–1265.

Linville, Patricia W., "Self-Complexity and Affective Extremity: Don't Put All of Your Eggs in One Cognitive Basket," *Social Cognition*, 1985, *3* (1), 94–120.

_ , "Self-complexity as a cognitive buffer against stress-related illness and depression.," *Journal of Personality and Social Psychology*, 1987, *52* (4), 663.

Malmendier, Ulrike and Geoffrey Tate, "CEO Overconfidence and Corporate Investment," *The Journal of Finance*, 2005, *60* (6), 2661–2700.

Mas-Colell, Andreu, Michael Whinston, and Jerry Green, *Microeconomic Theory*, Oxford University Press, 1995.

Maslow, Abraham Harold, "The Dynamics of Psychological Security-Insecurity.," *Character & Personality; A Quarterly for Psychodiagnostic & Allied Studies*, 1942.

Matt, Georg E., Carmelo Vázquez, and W. Keith Campbell, "Mood-Congruent Recall of Affectively Toned Stimuli: A Meta-Analytic Review," *Clinical Psychology Review*, 1992, *12* (2), 227–255.

Mayer, John D., Laura J. McCormick, and Sara E. Strong, "Mood-Congruent Memory and Natural Mood: New Evidence," *Personality and Social Psychology Bulletin*, 1995, *21* (7), 736–746.

_ **, Yvonne N. Gaschke, Debra L. Braverman, and Temperance W Evans**, "Mood-Congruent Judgment is a General Effect.," *Journal of Personality and Social Psychology*, 1992, *63* (1), 119.

McGuigan, Martin, Sandra McNally, and Gill Wyness, "Student Awareness of Costs and Benefits of Educational Decisions: Effects of an Information Campaign," *Journal of Human Capital*, 2016, *10* (4), 482–519.

Morf, Carolyn C. and Frederick Rhodewalt, "Unraveling the Paradoxes of Narcissism: A Dynamic Self-Regulatory Processing Model," *Psychological Inquiry*, 2001, *12* (4), 177–196.

O'Donoghue, Ted and Matthew Rabin, "Doing It Now or Later," *American Economic Review*, 1999, *89* (1), 103–124.

Orphanides, Athanasios and David Zervos, "Rational Addiction With Learning and Regret," *Journal of Political Economy*, 1995, *103* (4), 739–758.

Oster, Emily, Ira Shoulson, and Earl R. Dorsey, "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease," *American Economic Review*, 2013, *103* (2), 804–30.

Pelham, Brett W., "Self-Investment and Self-Esteem: Evidence for a Jamesian model of Self-Worth," *Journal of Personality and Social Psychology*, 1995, *69* (6), 1141.

Raskin, Robert and Howard Terry, "A Principal-Components Analysis of the Narcissistic Personality Inventory and Further Evidence of its Construct Validity.," *Journal of Personality and Social Psychology*, 1988, *54* (5), 890.

Sakulku, Jaruwan, "The Impostor Phenomenon," *The Journal of Behavioral Science*, 2011, *6* (1), 75–97.

Scheinkman, Jose A. and Wei Xiong, "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 2003, *111* (6), 1183–1220.

Schelling, Thomas C, "The Mind as a Consuming Organ," *The Multiple Self*, 1987, pp. 177–96.

Solomon, Andrew, *The Noonday Demon: An Atlas of Depression*, Simon and Schuster, 2014.

Spiegler, Ran, "Bayesian Networks and Boundedly Rational Expectations," *Quarterly Journal of Economics*, 2016, *131* (3), 1243–1290.

Spinnewijn, Johannes, "Unemployed But Optimistic: Optimal Insurance Design With Biased Beliefs," *Journal of the European Economic Association*, 2015, *13* (1), 130–167.

Turner, R. Jay, Blair Wheaton, and Donald A. Lloyd, "The Epidemiology of Social Stress," *American Sociological Review*, 1995, pp. 104–125.

Uluduz, Hatice and Ilhan Gunbayi, "Growth Mindset in the Classroom," *European Journal of Education Studies*, 2018.

Urdan, Tim and Carol Midgley, "Academic Self-Handicapping: What we Know, What More There is to Learn," *Educational Psychology Review*, 2001, *13* (2), 115–138.

Watkins, Philip C., Andrew Mathews, Donald A. Williamson, and Richard D. Fuller, "Mood-Congruent Memory in Depression: Emotional Priming or Elaboration?," *Journal of Abnormal Psychology*, 1992, *101* (3), 581.

_ , Karen Vache, Steven P. Verney, and Andrew Mathews, "Unconscious Mood-Congruent Memory Bias in Depression.," *Journal of Abnormal Psychology*, 1996, *105* (1), 34.

Winkelmann, Liliana and Rainer Winkelmann, "Why are the Unemployed so Unhappy? Evidence from Panel Data," *Economica*, 1998, *65* (257), 1–15.

Young, Valerie, *The Secret Thoughts of Successful Women: Why Capable People Suffer from the Impostor Syndrome and How to Thrive in Spite of it*, Crown Pub, 2011.

Appendices

A Proofs

Proof of Fact 1.

Because $g(k(0-\tilde{a})) = g(k\tilde{a})$ by the assumption on $g(\cdot)$, $E[s_i|\tilde{a}] = \frac{g(k(1-\tilde{a}))a}{g(k(1-\tilde{a}))a+g(k(0-\tilde{a}))(1-a)} = \frac{g(k-k\tilde{a})a}{g(k-k\tilde{a})a+g(k\tilde{a})(1-a)}$. Note that for any binary database, each fixed database is represented by a single parameter $a \in [0,1]$. Let $G_a(\tilde{a}) \equiv E[s_i|\tilde{a}]$ for each fixed database with a.

We first show that $G_a(\tilde{a})$ is weakly increasing in $\tilde{a}$. By taking the derivative of $G_a(\tilde{a})$,

$$\begin{aligned} G'_a(\tilde{a}) &= \frac{-kg'(k-k\tilde{a})a[g(k-k\tilde{a})a+g(k\tilde{a})(1-a)] - g(k(1-\tilde{a}))a[-kg'(k-k\tilde{a})a+kg'(k\tilde{a})(1-a)]}{[g(k-k\tilde{a})a+g(k\tilde{a})(1-a)]^2} \\ &= \frac{-ka(1-a)[g'(k\tilde{a})g(k-k\tilde{a})+g'(k-k\tilde{a})g(k\tilde{a})]}{[g(k-k\tilde{a})a+g(k\tilde{a})(1-a)]^2}. \end{aligned}$$

As $g'(x) \leq 0$ for $x > 0$ by the assumption on $g(\cdot)$, $G'_a(\tilde{a}) \geq 0$.

Now we prove Fact 1. Because $G_a(0) \geq 0$, $G_a(1) \leq 1$, and $G_a(\tilde{a})$ is continuous and increasing in $\tilde{a}$, there exists $\tilde{a}^* \in [0,1]$ such that $G_a(\tilde{a}^*) = \tilde{a}^*$ and $G'_a(\tilde{a}^*) \leq 1$. We show that for almost any database, $\tilde{a}^*$ is a SEP. Consider an equation $f(\tilde{a}^*;a) \equiv G_a(\tilde{a}^*) - \tilde{a}^* = 0$. Because $\frac{\partial f(\tilde{a}^*;a)}{\partial a} = \frac{g(k\tilde{a})g(k-k\tilde{a})}{[g(k-k\tilde{a})a+g(k\tilde{a})(1-a)]^2} \neq 0$, $Df(\tilde{a}^*;a) = \left(\frac{\partial f(\tilde{a}^*;a)}{\partial \tilde{a}^*}, \frac{\partial f(\tilde{a}^*;a)}{\partial a}\right)$ has rank 1. Then, by the Transversality Theorem, $\frac{\partial f(\tilde{a}^*;a)}{\partial \tilde{a}^*} = G'_a(\tilde{a}^*) - 1 \neq 0$ for almost any database a.[10] Because $G'_a(\tilde{a}^*) - 1 \leq 0$, we have shown that $G'_a(\tilde{a}^*) < 1$ for almost any database a. □

Proof of Proposition 1.

As derived in Fact 1, $E(s_i|\tilde{a}) = G_a(\tilde{a}) = \frac{g(k-k\tilde{a})a}{g(k-k\tilde{a})a+g(k\tilde{a})(1-a)}$. We show that $G_a(\tilde{a})$ has at most one inflection point, which ensures that there are at most two SEPs.

As shown in Fact 1, $G'_a(\tilde{a}) \geq 0$. Also,

$$\begin{aligned} G''_a(\tilde{a}) =& \frac{k^2a(1-a)}{[g(k-k\tilde{a})a+g(k\tilde{a})(1-a)]^3} \\ & \cdot \big\{2[g'(k\tilde{a})g(k-k\tilde{a})+g'(k-k\tilde{a})g(k\tilde{a})] \cdot [-g'(k-k\tilde{a})a+g'(k\tilde{a})(1-a)] \\ & - [g''(k\tilde{a})g(k-k\tilde{a})-g''(k-k\tilde{a})g(k\tilde{a})] \cdot [g(k-k\tilde{a})a+g(k\tilde{a})(1-a)]\big\}. \end{aligned}$$

[10] See, for example, Proposition 17.D.3 in Mas-Colell et al. (1995).

If the derivative of the terms in the curly bracket of $G_a''(\tilde{a})$ is always non-positive, then $G_a''(\tilde{a})$ crosses zero at most once, and hence, $G_a(\tilde{a})$ has at most one inflection point. Note that the derivative is:

$$\begin{aligned}&-k[g'''(k\tilde{a})g(k-k\tilde{a})+g'''(k-k\tilde{a})g(k\tilde{a})]\cdot[g(k-k\tilde{a})a+g(k\tilde{a})(1-a)]\\&+3k[g'(k\tilde{a})g(k-k\tilde{a})+g'(k-k\tilde{a})g(k\tilde{a})]\cdot[g''(k-k\tilde{a})a+g''(k\tilde{a})(1-a)].\end{aligned}\tag{1}$$

As $g(x)\geq 0$ and $g'(x)\leq 0$, (1) is non-positive when $g''(x)\geq g(x)(\geq 0)$ and $g'''(x)\geq 0$ for all $x>0$. Now suppose that $g'''(x)<0$ for some $x>0$. Even in this case, if $g'(x)\leq g'''(x)(<0)$ and $g''(x)\geq g(x)(\geq 0)$ hold for all $x>0$, then (1) is non-positive. □

Proof of Proposition 2.

I. By definition, $G_0(\tilde{a})=0$ and $G_1(\tilde{a})=1$ for any $k>0$ and $\tilde{a}\in[0,1]$. □

II. Because $G_a'(\tilde{a})=\frac{-ka(1-a)[g'(k\tilde{a})g(k-k\tilde{a})+g'(k-k\tilde{a})g(k\tilde{a})]}{[g(k-k\tilde{a})a+g(k\tilde{a})(1-a)]^2}$, there exists $\underline{k}>0$ such that if $k<\underline{k}$, then $G_a'(\tilde{a})<1$ for all $\tilde{a}\in[0,1]$. This ensures that the agent does not have multiple SEPs. □

III. Because of the assumption that a solution to the equation $\tilde{a}=E[s_i|\tilde{a}]$ is always locally unique, it suffices to show that there exists an unstable fixed point $G_a(\tilde{a})=\tilde{a}\in(0,1)$ for any sufficiently large $k>0$.

Because of the assumption that a solution to the equation $\tilde{a}=E[s_i|\tilde{a}]$ is always locally unique, $G_a(0)\geq 0$, and $G_a(1)\leq 1$, such an unstable fixed point $\tilde{a}$ exists if for any sufficiently large $k>0$ there exist $\tilde{a}_L,\tilde{a}_H\in(0,1)$ such that $\tilde{a}_L<\tilde{a}_H$, $G_a(\tilde{a}_L)<\tilde{a}_L$, and $G_a(\tilde{a}_H)>\tilde{a}_H$ (i.e., $G_a(\tilde{a})$ crosses the 45-degree line from below at $\tilde{a}\in(\tilde{a}_L,\tilde{a}_H)$). To show the existence of such $\tilde{a}_L$, take $\tilde{a}_L=\frac{1}{k}$ for any $k>2$. Because $G_a(\tilde{a}_L)=\frac{g(k-k\tilde{a}_L)a}{g(k-k\tilde{a}_L)a+g(k\tilde{a}_L)(1-a)}=\frac{a}{a+\frac{g(1)}{g(k-1)}(1-a)}$, $\frac{G_a(\tilde{a}_L)}{\tilde{a}_L}=\frac{a}{\frac{a}{k}+\frac{g(1)}{g(k-1)k}(1-a)}$. Because the right hand side converges to zero as k approaches $+\infty$, we know that for a sufficiently large k, there exists $\tilde{a}_L=\frac{1}{k}$ such that $\frac{G_a(\tilde{a}_L)}{\tilde{a}_L}<1$.

For the later use, we prove a slightly general version for $\tilde{a}_H$: for any $\epsilon\in(0,1]$, there is $\bar{k}$ such that for any $k>\bar{k}$ a high SEP $\tilde{a}^{high}$ is larger than $\frac{k-\epsilon}{k}$. Take $\tilde{a}_H=\frac{k-\epsilon}{k}$. Because $G_a(\tilde{a}_H)=\frac{g(k-k\tilde{a}_H)a}{g(k-k\tilde{a}_H)a+g(k\tilde{a}_H)(1-a)}=\frac{g(\epsilon)a}{g(\epsilon)a+g(k-\epsilon)(1-a)}$, $G_a(\tilde{a}_H)-\tilde{a}_H=\frac{g(\epsilon)a}{g(\epsilon)a+g(k-\epsilon)(1-a)}-\frac{k-\epsilon}{k}=\frac{g(\epsilon)a\epsilon-(k-\epsilon)g(k-\epsilon)(1-a)}{k[g(\epsilon)a+g(k-\epsilon)(1-a)]}$. Because the numerator is positive as k $to+\infty$, we know that for a sufficiently large k, there exists $\tilde{a}_H=\frac{k-\epsilon}{k}$ such that $G_a(\tilde{a}_H)>\tilde{a}_H$.

Hence, for any $a \in (0,1)$, there is a $\overline{k} > 0$ such that if $k > \overline{k}$, then the agent has fragile self-esteem. □

Proof of Proposition 3.

I. As $G_a(a) = \frac{g(k-ka)a}{g(k-ka)a+g(ka)(1-a)}$, for any $a \in (0,1)$, $G_a(a) > a$ if $a > 1/2$; $G_a(a) = a$ if $a = 1/2$; and $G_a(a) < a$ if $a < 1/2$. Suppose the case in which $a > 1/2$ and hence $G_a(a) > a$. Because $G_a(1) \leq 1$, there exists $\tilde{a} \in (a,1)$ such that $G_a(\tilde{a}) = \tilde{a}$. For a sufficiently small k, this is the unique SEP by Proposition 2 Part II. The cases in which $a = 1/2$ and $a < 1/2$ can be shown in the same manner. □

II. By Proposition 2 Part III, there are two SEPs in this case. Moreover, one of these SEPs is smaller than $\tilde{a}_L$ and another is larger than $\tilde{a}_H$, where $\tilde{a}_L$ and $\tilde{a}_H$ are specified in the proof of Proposition 2 Part III. Thus, if we take $k > \max\{a, \frac{1}{1-a}, \overline{k}\}$ where $\overline{k}$ is specified in Proposition 2 Part III, one SEP is strictly smaller than a while another SEP is strictly larger than a. □

III. It is straightforward to show that $\lim_{k\to 0} G_a(\tilde{a}) = a$. In the proof of Proposition 2 Part III, we showed that for a sufficiently large k, there are two SEPs where one SEP is smaller than $\frac{1}{k}$ and another SEP is larger than $\frac{k-1}{k}$. Taking $k \to \infty$ completes the proof. □

Proof of Proposition 4.

Suppose there are two SEPs. Because of the assumption that a solution to the equation $\tilde{a} = E[s_i|\tilde{a}]$ is always locally unique, $G_a(0) \geq 0$, and $G_a(1) \leq 1$, there is one unstable fixed point $\tilde{a}^u = E[s_i|\tilde{a}^u] \in (0,1)$. Note that a small increase in a makes the high SEP more stable and the low SEP less stable if $\frac{d\tilde{a}^u}{da} < 0$.

Let denote $f(\tilde{a}^u, a) = E[s_i|\tilde{a}^u] - \tilde{a}^u = 0$. Note that we have $\frac{\partial f}{\partial \tilde{a}^u} = G'_a(\tilde{a}^u) - 1 > 0$. Hence, by the implicit function theorem, we obtain

$$\begin{aligned}
\frac{d\tilde{a}^u}{da} &= -\frac{\partial f/\partial a}{\partial f/\partial \tilde{a}^u} \\
&= -\frac{1}{G'_a(\tilde{a}^u) - 1} \cdot \frac{g(k-k\tilde{a}^u)[g(k-k\tilde{a}^u)a + g(k\tilde{a}^u)(1-a)] - g(k-k\tilde{a}^u)a[g(k-k\tilde{a}^u) - g(k\tilde{a}^u)]}{[g(k-k\tilde{a}^u)a + g(k\tilde{a}^u)(1-a)]^2} \\
&= -\frac{1}{G'_a(\tilde{a}^u) - 1} \cdot \frac{g(k-k\tilde{a}^u)g(k\tilde{a}^u)}{[g(k-k\tilde{a}^u)a + g(k\tilde{a}^u)(1-a)]^2} < 0. \quad \square
\end{aligned}$$

Proof of Proposition 5.

I. In what follows, we show that $k\frac{d\tilde{a}}{dk}$ converges to zero as $k \to 0$. Combined with Proposition 3 Part III, it implies that $\lim_{k\to 0} \frac{dk\tilde{a}}{dk} = \lim_{k\to 0} \tilde{a} + \lim_{k\to 0} k\frac{d\tilde{a}}{dk} = a > 0$, and hence, there exists a sufficiently small $\underline{k} > 0$ such that $\frac{dk\tilde{a}}{dk} > 0$ for any $k < \underline{k}$.

Let denote $f(\tilde{a}, k) = E[s_i|\tilde{a}] - \tilde{a} = 0$. Note that we have $\frac{\partial f}{\partial \tilde{a}} = G'_a(\tilde{a}) - 1 < 0$ by the definition of SEP. Hence, by the implicit function theorem, we obtain

$$\begin{aligned}
k\frac{d\tilde{a}}{dk} =& -k\frac{\partial f/\partial k}{\partial f/\partial \tilde{a}} \\
=& -k\frac{1}{G'_a(\tilde{a}) - 1} \cdot \frac{a(1-\tilde{a})g'(k-k\tilde{a})[g(k-k\tilde{a})a + g(k\tilde{a})(1-a)] - ag(k-k\tilde{a})[a(1-\tilde{a})g'(k-k\tilde{a}) + (1-a)\tilde{a}g(k\tilde{a})]}{[g(k-k\tilde{a})a + g(k\tilde{a})(1-a)]^2} \\
=& -k\frac{[g(k-k\tilde{a})a + g(k\tilde{a})(1-a)]^2}{-ka(1-a)[g'(k\tilde{a})g(k-k\tilde{a}) + g'(k-k\tilde{a})g(k\tilde{a})] - [g(k-k\tilde{a})a + g(k\tilde{a})(1-a)]^2} \\
& \cdot \frac{a(1-a)[g'(k-k\tilde{a})g(k\tilde{a})(1-\tilde{a}) - g(k-k\tilde{a})g'(k\tilde{a})\tilde{a}]}{[g(k-k\tilde{a})a + g(k\tilde{a})(1-a)]^2} \\
=& -\frac{a(1-a)[g'(k-k\tilde{a})g(k\tilde{a})(k-k\tilde{a}) - g(k-k\tilde{a})g'(k\tilde{a})k\tilde{a}]}{-ka(1-a)[g'(k\tilde{a})g(k-k\tilde{a}) + g'(k-k\tilde{a})g(k\tilde{a})] - [g(k-k\tilde{a})a + g(k\tilde{a})(1-a)]^2}.
\end{aligned}$$

Note that the numerator converges to zero as $k \to 0$, whereas the denominator converges to $-g(0)^2 < 0$ as $k \to 0$. Hence, $\lim_{k\to 0} k\frac{d\tilde{a}}{dk} = 0$. □

II. Suppose Example 1 with different levels of k. As in the main text, if $k > 2$, there are two SEPs in which $\tilde{a}^{low} = \frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{k^2}}\right)$ and $\tilde{a}^{high} = \frac{1}{2}\left(1 + \sqrt{1 - \frac{4}{k^2}}\right)$. Note that $k\tilde{a}^{low} = \frac{1}{2}\left(k - \sqrt{k^2 - 4}\right)$. By taking the derivative with respect to $k > 2$, we obtain:

$$\frac{dk\tilde{a}^{low}}{dk} = \frac{1}{2}\left(1 - \frac{k}{\sqrt{k^2-4}}\right) = \left(1 - \frac{1}{\sqrt{1 - \frac{4}{k^2}}}\right) < 0.$$

Because the agent's effort is strictly increasing in $k\tilde{a}^{low}$, the agent lowers his effort for a higher $k > 2$ in this case. □

III. In the proof of Proposition 2 Part III, we showed that for a sufficiently large k, a high SEP $\tilde{a}^{high}$ is larger than $\frac{k-\epsilon}{k}$ for any $\epsilon \in (0, 1]$.

In what follows, we show that $k\frac{d\tilde{a}^{high}}{dk}$ is non-negative for sufficiently large $k > 0$. It implies that $\lim_{k\to\infty} \frac{dk\tilde{a}^{high}}{dk} = \lim_{k\to\infty} \tilde{a}^{high} + \lim_{k\to\infty} k\frac{d\tilde{a}^{high}}{dk} > 0$, and hence, there exists a sufficiently large $\bar{k} > 0$ such that $\frac{dk\tilde{a}^{high}}{dk} > 0$ for any $k > \bar{k}$.

Let denote $f(\tilde{a}^{high}, k) = E[s_i|\tilde{a}^{high}] - \tilde{a}^{high} = 0$. Note that we have $\frac{\partial f}{\partial \tilde{a}^{high}} = G'_a(\tilde{a}^{high}) - 1 < 0$ by the definition of SEP. As in Part I of this proposition, by the implicit function theorem, we obtain

$$k\frac{d\tilde{a}^{high}}{dk} = -\frac{1}{G'_a(\tilde{a}^{high}) - 1} \cdot \frac{a(1-a)[g'(k-k\tilde{a}^{high})g(k\tilde{a}^{high})(k-k\tilde{a}^{high}) - g(k-k\tilde{a}^{high})g'(k\tilde{a}^{high})k\tilde{a}^{high}]}{[g(k-k\tilde{a}^{high})a + g(k\tilde{a}^{high})(1-a)]^2}.$$

Note that $k\frac{d\tilde{a}^{high}}{dk} \geq 0$ if and only if the numerator is non-negative.

Because $\frac{dg(x)x}{dx} = g'(x)x + g(x)$ and $k\tilde{a}^{high} > k \cdot \frac{k-1}{k} = k - 1$, the assumption implies that $g(k\tilde{a}^{high}) \leq -g'(k\tilde{a}^{high})k\tilde{a}^{high}$ for sufficiently large k. Then:

$$\begin{aligned}
&g'(k-k\tilde{a}^{high})g(k\tilde{a}^{high})(k-k\tilde{a}^{high}) - g(k-k\tilde{a}^{high})g'(k\tilde{a}^{high})k\tilde{a}^{high} \\
\geq &g'(k-k\tilde{a}^{high})g(k\tilde{a}^{high})(k-k\tilde{a}^{high}) + g(k-k\tilde{a}^{high})g(k\tilde{a}^{high}) \\
= &g(k\tilde{a}^{high})[g'(k-k\tilde{a}^{high})(k-k\tilde{a}^{high}) + g(k-k\tilde{a}^{high})]. \qquad (2)
\end{aligned}$$

Note that $\frac{dg(x)x}{dx} = g'(x)x + g(x)$ is increasing for sufficiently small $\epsilon > 0$. Because $k - k\tilde{a}^{high} < k - k \cdot \frac{k-\epsilon}{k} = \epsilon$, for any $g(\cdot)$, by taking sufficiently large k, we can obtain $\epsilon > 0$ such that $\frac{dg(x)x}{dx} \geq 0$ for any $x \in [0, \epsilon)$. Hence, if k is sufficiently large, (2) is non-negative. □

Proof of Proposition 6.

I. If the agent is not fragile, from any seed self-view, his resulting level of SEP is the same. If $\delta = 0$, his action cannot affect anything, and hence, he choooses $e = 0$. □

II. (i) Suppose the situation in which $\tilde{a} = \epsilon$ is just below the unstable fixed point. Because exerting the effort discontinuously increases his resulting level of self-esteem whereas $c'(0) = 0$, he chooses $e > 0$ in such a case. Hence, his average action is positive.

(ii) Given $\tilde{a}_{-1} + \epsilon$, the agent chooses either $e = 0$ or $e = \frac{\tilde{a}^u - \tilde{a}_{-1} - \epsilon}{\delta}$ where $\tilde{a}^u$ is an unstable fixed point. Note that $\frac{d\tilde{a}^u}{da} < 0$ as in the proof of Proposition 4 Part I. Hence, given $\tilde{a}_{-1}$ and ϵ, as a increases, the agent either continuously strictly decreases e or starts exerting effort (i.e., discontinuously increases e from 0 to $\frac{\tilde{a}^u - \tilde{a}_{-1} - \epsilon}{\delta}$). □