Ego Utility and Information Acquisition[1]

Botond Kőszegi, UC Berkeley

Abstract

Based on extensive psychological evidence and the experience of most of us, it seems obvious that people intrinsically care about the perceptions of themselves, not only because it helps in making decisions. This paper explores some of the consequences of this motivation in a model where the agent derives utility from both financial outcomes and her beliefs about her ability ('ego utility'). The model can explain a variety of anomalous-looking behavior, like refusing to consider new information about past judgments, putting off making judgments, and holding on too long to losing decisions. Several applications are discussed, with particular attention to how to provide incentives to employees with ego utility.

[1]First Version: July 1999. A version of this paper also appeared in my MIT Ph.D. dissertation (June, 2000).

1 Introduction

When it comes to being honest about ourselves, most of us have little to be proud of. We tend to attribute our successes to skill and our failures to bad luck, to rate our performance in tasks better than observers do and than objective criteria would warrant, to believe too strongly in what are merely first impressions, and to avoid dealing with our mistakes.

While these facts are obvious to both laypeople and psychologists, they are only now starting to be incorporated into economic models. The above views about the self can arise for strategic reasons in a setting with time-inconsistency (Carillo and Mariotti 1997, Carillo 1997, Benabou and Tirole 1999a, Benabou and Tirole 1999b), as the decisionmaker tries to use her beliefs to influence her future selves. In these models, agents are rational and have standard preferences (besides being time-inconsistent). Others (Rabin and Schrag 1999, Gervais and Odean 1999) assume that distorted views are simply a result of cognitive mistakes, and derive the implications of these mistakes. Here, agents are irrational–they don't realize they are making a mistake. The current paper takes a third approach, putting these phenomena on an emotional instead of a strategic or cognitive level. I simply assume that people care about their perceptions of themselves, and consequently actively manage their self-image. Formally, there is a non-standard dimension of utility, called 'ego utility,' which is a function of the agent's beliefs about her ability. Ego utility enters additively besides a more standard, instrumental side of utility, that derived from financial payoffs. The agent can sample signals about the payoffs of different financial options, but in doing so she indirectly also acquires information about her ability: higher types receive more accurate signals–which I interpret as subjective judgments,–so any feedback about previous judgments is informative about the agent's type. But the agent can, to some extent, manipulate her beliefs about herself by choosing not to acquire useful information about her available choices. I assume that the agent's ego utility function is concave, that is, she is 'information averse'– she likes good news about her ability less than she dislikes bad news. This seems to be

both the more realistic, and, in the paper's context, where the agent gathers free useful information, the more interesting assumption. In other respects, the model is a neoclassical one: the agent is a Bayesian expected utility maximizer.

The model's simple setup implies that the agent is averse to the combination of making a subjective judgment *and* reviewing it later, since the two together are informative about ability. This has three possible consequences. First, having made a subjective judgment, the agent is reluctant to review it later, since the review might put her earlier judgment in a bad light. Therefore, the agent is *sluggish* in responding to new information available to her. The flip side of this effect is that if the agent expects to have to review her judgments later, she will be reluctant to make them in the first place; in other words, she *procrastinates* in making up her mind about the choices. And finally, information-averse agents are averse to making subjective judgments as such–they might delegate the responsibility of making assessments to a non-subjective source, even one of questionable quality.

These effects depend in interesting ways on the decisionmaker and the environment. Section 2.4 considers two examples: the confidence of the decisionmaker (her prior probability of being a high type) and an environment with feedback. A more confident agent is less likely to procrastinate, since she considers her initial signal to be more valuable. However, she will tend to be more sluggish: having made a judgment, she doesn't feel the need to reconsider it later. Similarly, feedback is a mixed blessing in terms of the quality of the agent's financial decisions. If she expects involuntary feedback on her previous judgment that she could voluntarily get now, she won't avoid the current information so much. That is, she is less sluggish. However, forced feedback can undermine the agent's willingness to make a subjective judgment to start with–she doesn't want her judgment to be proven wrong by the unwanted information.

The models in this paper are intended to fit a number of economic applications. In section 5, three applications, stock market participation, small businesses, and project choice by managers are considered in the light of the current models. A notable set of applications

for these models is to employee motivation, and this application is discussed in some detail. If intrinsic motivation is understood as a desire to manage one's self-image while performing one's job, performance bonuses might discourage information-averse workers because they provide information about ability. This puts a twist on the standard principal-agent framework, resulting in crucially different optimal incentives. Employers might deliberately want to condition pay on a noisy signal of performance and ability, thereby drowning the inferences the employee can make about herself from her pay. Depending on the worker's risk aversion, the employer might want to condition pay a lot on a very noisy signal of ability and effort, or not use incentives at all. The surprising result is that if the worker's job is information-sensitive (she needs to make decisions based on subjective judgments), then, in determining the optimal incentives, the strength of the employee's information-aversion can be more important than her risk aversion: she might receive noisy incentives, no matter how risk-averse she is. The intuition is that it is not enough to compensate the employee for the 'ego pain' inflicted on her, she has to be induced to actually base her decisions on the appropriate judgment. This is expensive to do if the signal is less noisy.

This paper is closely related to Kőszegi (2000), which argues extensively that the ego utility approach is the most sensible model of many phenomena related to self-image, as well as demonstrating that information-aversion is likely to be quite common with financial stakes present. Just as this paper, Kőszegi (2000) posits that agents derive ego utility from their beliefs about themselves. Specifically, they like to think that they are capable enough to perform the more ambitious of two activities, one in which only higher types can perform well. And once again, agents can influence the flow of information that they receive about themselves, giving them a degree of control over what they can believe about themselves. The paper shows that this is sufficient to produce overconfidence in beliefs: too many agents will honestly (and rationally) believe that they can perform the ambitious activity well. The paper also examines other logical consequences of this setup; for example, if the more ambitious of the two activities is also more informative, the overconfidence in beliefs might

not be expressed in the agent's actions–although she justly believes she would make more money with the ambitious option, she chooses the unambitious one for fear of finding out otherwise.

The current paper takes ego utility and information aversion as given, and in that sense builds on Kőszegi (2000). But it addresses a different question: how these preferences influence information acquisition about different options available to the decisionmaker. In the other paper, ability is simply taken to be a parameter that directly affects the agent's financial outcomes–why that parameter should influence success was not modeled. In addition, information gathering is limited to information about one's ability, whereas in the real world we gather a lot of information about the options we are about to face; in fact, most of our information about the self derives indirectly from our performance in specific skill-sensitive tasks.

2 Ego Utility and Financial Judgments

Before people make important decisions that affect their life outcomes, they usually have to make judgments about the relative merits of options they are about to face. Some people are better at making these judgments than others, and it seems introspectively and observationally obvious that people like to see themselves as capable of making them well [2]. If this is the case, information is not only collected to improve decisions, but also to manage one's perception of the self. The current section models decision-making consequences of this

[2] For both direct and indirect psychological evidence on this point, see Kőszegi (2000). The direct evidence is taken from the cognitive dissonance literature, which indicates that negative judgments about the self make people very uncomfortable. For example, those who (at the experimenter's request) call an unknown person stupid feel bad about it, so much so that they convince themselves that the other person *is* actually stupid. The indirect evidence comes from the fact that people hold incredibly positive views about themselves: in anything from employment through taking credit or blame for outcomes to prospects for living a healthy life, the majority of us think that we are in a better shape than the median person.

motivation. I will use a general language to discuss the model; specific applications are considered in section 5. One useful example to keep in mind for intuition is that of stock market participation: participating involves choosing between stocks, constantly making judgments about their relative merits.

2.1 Basic setup

The setup of the choice problem is the following. The agent has to choose one of two options (stocks) in each of two consecutive periods. Option 1 is riskless with the return $a_1 \in (0,1)$. Option 2 is risky with $a_2 \in \{0,1\}$, but it pays off the same amount in both periods. (This setup is not inconsistent with the motivating example of choosing stocks: it is equivalent to the agent just getting signals about the *difference* in returns of the options.) Judgments are modeled as signals s^t that can be voluntarily observed about a_2 in period t before the decision has to be made. The ability to observe s^2 is not tied to having chosen option 2 in period 1. The exact timing of the problem is the following:

1: signal s^1 (choose to either observe it or not)
1': choice (payoff not observed)
2: signal s^2 (choose to either observe it or not)
2': choice (payoff not observed)
2": ego utility realized
3: financial outcomes realized

As an example, the following choice problem has a time structure resembling the above. The agent gets an opportunity to learn about and invest in a firm which will eventually

succeed or fail. Later, when new information about the firm is available, the agent can once again decide whether to invest. The actual financial outcome is further down the line. Choosing to observe the signal s^1 or s^2 corresponds to making a judgment or reviewing a judgment about the firm, respectively [3]. After possibly reviewing the options and choosing one of them, the agent has to confront how what she has seen reflects on her ability; then, her utility from self-image (ego utility) is realized. Since in the present paper we are primarily concerned with information gathering for choice, we assume that the financial outcomes are realized so late as not to affect the ego; if the information implicit in the observation of financial payoffs also affected the ego, the discussion would be cluttered by many additional cases and effects.

The signal s^t is imperfectly correlated with the actual payoff of option 2. In particular, the space of the signals is also $\{0,1\}$, and the probability that one is 'right' varies with the agent's type and the nature of the signal. We distinguish between two kinds of signals. The probability that a *type-dependent* signal in period t is right can be one of two values:

$$Prob(s^t = a_2) = q^t = \begin{cases} q_H^t & \text{if agent is high-type;} \\ q_L^t & \text{if agent is low-type.} \end{cases} \tag{1}$$

We assume $q_H^t > q_L^t \geq \frac{1}{2}$. First-period signals are always type-dependent to capture the notion that early decisions, when things are usually not so clear yet, depend more on a subjective judgment. In contrast, for the second period, we will consider both the case of type-dependent and type-independent signals. The latter kind of signal is accurate with probability $q_I > \frac{1}{2}$, independently of the agent's type. The agent's priors are summarized in $p_{jk}^0 = Prob(a_2 = j, q^t = q_k^t)$, and I use the notation $p_{jk}(S)$ for the agent's posteriors after

[3]Modeling judgments as signals is a simplification. A more realistic view is that the agent collects decision-relevant information, and the real judgment she has to make involves deciphering this information, for example by choosing relative weights of importance. It is probably the latter step that better investors can make better. In my formulation, the mental process of making a judgment is collapsed into a reduced form.

observing the set of signals S. Let S^t be the set of signals observed by the end of period t.

Utility from self-image in this problem depends on the (Bayesian) agent's subjective probability of being a high type. At the end of period 2, this probability is given by $p_{1H}(S^2)+p_{0H}(S^2)$. Total utility is then

$$wu\left(p_{1H}(S^2)+p_{0H}(S^2)\right)+n_1a_1+(2-n_1)a_2, \tag{2}$$

where n_1 is the number of times the agent chooses option 1. The agent is a Bayesian expected utility maximizer; see (Kőszegi 2000) for a justification of this kind of model.

$w>0$ is simply a weighting parameter. As noted before, this section focuses on an information-averse decisionmaker, so we assume that u is strictly concave. There are two reasons to do this. First, I argued in Kőszegi (2000) that–with financial stakes present–self-image protection or information-aversion is likely to be the more common phenomenon affecting agents with ego utility [4]. In addition, in a model of this type, self-image protection is more interesting: it counterbalances the classical value of information.

2.2 Preliminary results

We are primarily interested in what kinds of signals or combinations of signals are informative about the agent's type, since this is what the interesting effects will depend on.

We start with two obvious facts.

Fact 1 *Any combination of type-independent signals is uninformative about ability.*

Fact 2 *(Risk neutrality) In each period $t=1,2$, the agent chooses option 1 if and only if*

$$a_1>p_{1H}(S^t)+p_{1L}(S^t), \tag{3}$$

[4] The main reason for this is that in many situations, there are likely to be financially less costly ways for agents to learn about themselves. Thus, if they are information-loving, people are more likely to use those channels to find out things about themselves. For self-image protection, however, there is no way to substitute to a less costly channel.

That is, the agent maximizes the expected return conditional on her information: she chooses option 1 if and only if return is higher than the posterior probability that option 2 would yield an outcome of 1.

The following lemma is a key intermediate result. Its proof is relatively straight-forward, but requires a few steps, so it is relegated to the appendix.

Lemma 1 *A type-dependent signal combined with any other signal is always informative about ability.*

Although it is not quite accurate in general, the intuition for the case when $p^0_{0H} = p^0_{1H}, p^0_{0L} = p^0_{1L}$ is the most useful to understand. In that case, a type-dependent signal by itself is not informative–since both outcomes for option 2 are equally likely a priori, making a judgment either way doesn't say anything about the person. Then, since high types are more likely to receive 'correct' signals, receiving consistent informative signals, or, in plainer terms, 'not getting confused,' is a sign of being a good decisionmaker. And since any two signals are either consistent or inconsistent, the two signals are informative [5].

2.3 Sluggishness and Procrastination

We concentrate on the case of independent and neutral priors. Let $r_2 = Prob(a_2 = 1) = \frac{1}{2}$, and assume that the prior probability c of being a high type, the trait relevant for ego utility, is independent of a_2. This probability will be interpreted as confidence; thus the notation c. We assume that $0 < c < 1$, so that the agent is not completely certain of her ability.

[5] Using the proof of lemma 1, it is easy to show that *if* the type-dependent signal is uninformative by itself, then the good news about ability is if the two signals are the same. So in that case the intuition is still correct. It is not correct in general, though: if, for example, the agent is sure that $a_2 = 1$ and both signals are type-dependent, then it is better to receive a zero signal and a one signal than receiving two zero signals.

Note again that due to the timing of the problem, observing the financial outcome does not convey information that enters ego utility.

The next two theorems constitute the main results of section 2. They show, respectively, that if ego utility is important enough for the agent, then she will either fail to reconsider her choices, leading to a sluggishness in them, fail to make up her own mind about it at the first given opportunity (theorem 1), or, in certain conditions, wait for a type-independent signal to make up her mind (theorem 2).

Theorem 1 Sluggishness and Procrastination *Suppose that s^1 is type-dependent. If her ego utility is sufficiently important (w is sufficiently large), the agent will observe exactly one of the signals s^1 and s^2.*

Proof. That the agent won't observe both signals for a sufficiently large w is an obvious consequence of lemma 1. To show that one signal will in fact be observed, we show that a single signal is not informative about ability. For type-independent signals, this is implied by fact 1. For type-dependent signals, it follows from neutral priors: since $a_2 = 0$ and $a_2 = 1$ are equally likely, both types of agents receive the signal $s^1 = 1$ with probability $\frac{1}{2}$. Formally, we have

$$Prob(q^t = q_H^t | s^t = 1) = \frac{\frac{1}{2}cq_H^t + \frac{1}{2}c(1-q_H^t)}{\frac{1}{2}cq_H^t + \frac{1}{2}c(1-q_H^t) + \frac{1}{2}(1-c)q_L^t + \frac{1}{2}(1-c)(1-q_L^t)} = c. \quad (4)$$

This completes the proof. □

The following corollary illustrates the use of this theorem for $a_1 > \frac{1}{2}$, a case when the agent's prior beliefs favor option 1.

Corollary 1 *Suppose that $a_1 > \frac{1}{2}$ and that the first-period signal is type-dependent and the second one is type-independent. If ego utility is sufficiently important (w is sufficiently large), only one of the signals will be observed. It will be the second one if and only if*

$$2(cq_H^1 + (1-c)q_L^1 - a_1)_+ < (q_I - a_1)_+. \quad (5)$$

Proof. Appendix.

It's easy to generate decision rules for the other cases of the problem, but those are not worth writing down for our purposes.

This theorem summarizes two basic behavioral distortions that can arise as a consequence of ego utility. The first one I have labeled sluggishness: once the agent has made a judgment whose accuracy depends on her ability, she will be reluctant to look at new information later, afraid the new information would reveal the judgments she has made to be poor. Without new information, of course, the agent will choose the same option as before (see fact 2), exhibiting an excess sluggishness in changing options relative to the information available to her.

The second distortion to financial decisionmaking arises when the second-period signal is more instrumentally valuable than the first-period one. When the agent knows her choices will be evaluated in the future, she might not want to think about them seriously today, so that the later judgments are not reflective of her ability. That is, she puts off (procrastinates about) making a serious decision.

In the above setup, whether the information aversion of the agent played itself out in sluggishness or procrastination depended only on the informativeness of s^1 and s^2. But there is more to it than that. By fact 1, type-independent signals observed in isolation (not in combination with type-dependent signals) are not threatening to the ego, so agents have a general preference for type-independent signals irrespective of informativeness. Specifically, agents who expect to receive feedback about their choices later or to possibly reconsider them should have an incentive to wait for a type-independent signal even if it is not more accurate. To make this statement formally, one needs to expand the basic model a little bit. This is done in theorem 2.

Theorem 2 Deferral to objective criteria *Consider the same setup as in section 2.1*

with the following modifications. Now there are T periods of choice, with the same timing pattern as periods 1 and 2 in section 2.1, and the non-ego utilities are realized in period $T+1$. Suppose s^1 is type-dependent.

1. *If s^2 is type-independent, and there is a type-independent signal s^3 that has to be observed by the agent, then for a sufficiently large w only s^2, not s^1, will be observed.*

2. *If $s^2, s^3, \ldots, s^T$ are type-independent and s^1 is not perfectly informative ($c = q_H^1 = 1$), then for a sufficiently large w and a sufficiently large T and sufficiently informative $s^3, \ldots, s^T$ signals, s^2, and not s^1, will be observed.*

Proof. Immediate from fact 1 and lemma 1. □

There are other variants of the same principle. For example, the above theorem still holds true if s^2 is type-dependent, but does not distinguish the types well (for example, when q_H^2 and q_L^2 are close). Or, even with two periods, if the second-period signal is type-independent and in the first period, the agent can choose between a type-dependent and a type-independent signal, then for a sufficiently large w she will prefer the type-independent one even if it is less informative [6].

Procrastination in theorem 1 and theorem 2, then, describe different aspects of the unwillingness of agents to 'make up their minds' about choices that will have to be reviewed for some reason. One can defer this responsibility by relying on previous knowledge or tradition (priors) or by recruiting help that is not reflective of the self.

2.4 Confidence, Feedback, and Behavioral Distortions

Having reviewed the basic behavioral distortions due to self-image protection, it is natural to ask how these effects respond to different environments. We will consider two questions.

[6] Consider a group of hikers coming to a fork in the road, having little idea which way to go. If they can't figure out the likely way, many people in this situation prefer to flip a coin, even though that can't possibly increase the probability of making a good choice relative to a subjective judgment.

First, motivated by the paper's general focus on self-image, we examine how the agent's prior probability of being a high type (c) influences the two effects. This parameter is interpreted as confidence, and the posterior probability of being a high type is what ultimately determines ego utility. Second, we look at the consequences of giving the agent feedback about the options available to her. This is interesting because it is the economist's remedy for limited information gathering.

In the setup of corollary 1, the higher is the agent's prior probability of being a high type, the more likely it is that condition 26 is violated; that is, if it's violated for some c, it is also violated for $c' > c$. Thus we have the following.

Theorem 3 *Suppose s^1 is type-dependent and s^2 is type-independent. If the agent is sluggish for some $0 < c < 1$, she does not procrastinate for any $c' > c$.*

This theorem says that in some sense confidence helps in overcoming the procrastination problem. It is somewhat reminiscent of the effect found in a static setting by Weinberg (1999), where higher signals about ability lead the agent to take on more challenging tasks, which she would otherwise avoid.

The effect in theorem 3 is clearly driven by the fact that if c is higher, the first-period signal is perceived to be more informative by the agent, and so harder to give up. This intuition goes quite far. A first-period signal, at the very least, improves the first-period decision, and an agent who sees herself as a better decisionmaker will think it improves that decision more. Although there might be complications compared to the simple setup of theorem 3, this intuition is not reversed [7]. The only crucial caveat is that agents who are

[7] For example, if s^2 is also type-dependent, then a higher c also makes the second-period signal more informative. However, for this to outweigh the effect coming through the accuracy of the first-period signal, it has to be very strong (since the other signal is observed earlier). In particular, as long as $2(q_H^1 - q_L^1) > q_H^2 - q_L^2$, theorem 3 still holds true, with the proof being the same. And I would expect this relationship to hold in general: it is likely to require less skill to make a judgment later rather than earlier. (It might be the case

more certain of their ability will suffer less from *both* procrastination and sluggishness: they can learn less about themselves through making subjective judgments. In this setup, agents with confidence levels close enough to 0 or 1 won't avoid any signals. Again, an effect similar to this was also found in Weinberg (1999). So, to make an earlier statement more accurate, confidence helps in overcoming procrastination as long as it doesn't at the same time make the agent more unsure about her type.

But in a multi-period problem there is a flip side to the beneficial effects of confidence against procrastination. In contrast to what I have argued above for signals in the first period, it seems that signals in the second period might be less likely to be observed by confident agents. Consider, for example, $a_1 > \frac{1}{2}$, $s^1 = 1$, and s^2 type-dependent with equivalent informativeness to s^1; that is, $q_H^1 = q_H^2 = q_H$ and $q_L^1 = q_L^2 = q_L$. Further, make the assumption that $Prob(a_2 = 1|s^1 = 1) = cq_H + (1-c)q_L > a_1$. Then, the informativeness of the signal s^2 stems from the possibility that $s^2 = 0$, which leads to a reversal of the decision for any c: $Prob(a_2 = 1|s^1 = 1, s^2 = 0) = \frac{1}{2}$ for any c. So conditional on reversing the decision, the value of doing so is the same for all c. However, the probability (or the perceived probability) that the decision will be reversed is *decreasing* in c:

$$Prob(s^2 = 1|s^1 = 1) = c((1-q_H)^2 + q_H^2) + (1-c)((1-q_L)^2 + q_L^2), \tag{6}$$

which is increasing in c. Confident agents think that they can make good decisions the first time around, so they think it is less likely they would change their minds. Therefore, they don't care about reconsidering the decision too much.

This highlights a key distinction between early and late signals: while more confident agents always consider a first-period signal to be more informative than less confident agents, they might find a later signal less informative if they have already observed a signal earlier.

that information comes in between periods 1 and 2 that only high types can analyze, making $q_H^2 - q_L^2$ possibly higher than $q_H^1 - q_L^1$. However, in that case, the second-period decision is likely to be quite hard–the signal not being so informative,–so the first-period signal is observed for any c.)

This is the case when the crucial question is whether to reverse decisions. Thus, while confident people are likely to be less prone to procrastination, they are probably more prone to sluggishness.

Let us move on to the solution most commonly recommended by economists against ignorance: feedback, or, more generally, information. The problem is that agents sometimes don't want information because they want to protect their egos. Feedback, by disallowing them from getting caught up in an ignorance state, might be a remedy. Of course, the decisionmaker can in general avoid performance-relevant information at least for a while, no matter how hard it is forced on her [8]. Still, if our agent knows that she won't be able to 'flatter herself' in the long run, she might as well not flatter herself now. The logic is the same as in the case of an academic who likes to think her paper is really good and therefore tries not thinking about it too much, but if she has submitted it for publication and knows there is a judgment coming up about it soon, she will be more realistic. This intuition, related to the well-known fact in the psychology literature that self-serving biases diminish with the threat of verification (Fiske and Taylor 1991), is the core of the following very general theorem.

Theorem 4 *Take a T-period model as in theorem 2. Suppose some period i's signal s^i becomes available again in period $j > i$; that is, $s^j = s^i$. Then if the agent observes s^i if s^j doesn't have to be observed, she observes s^i when s^j has to be observed.*

Proof. In any state of nature in period i, consider the expected utility from following the optimal policy after observing versus not observing s^i. The expected utility when observing s^i is independent of whether s^j has to be observed, while not observing it yields a (weakly)

[8] It is an interesting, and clearly crucial, question on how one can provide feedback to someone who doesn't want to hear it. I'm not going to deal with this in any detail here.

lower expected utility if s^j has to be observed. Thus if s^i is observed when s^j doesn't have to be observed, it is also observed when s^j has to be observed. □

The converse is clearly not true, so forced feedback can undermine ignorance. This proof takes strong advantage of the fact that s^i and s^j are the same signal. How realistic is this assumption? There is at least one sense in which it is not: s^i could be a type-dependent signal, and it seems unreasonable to assume that anyone who might be providing the feedback would have access to that signal at any time. If the signals are not the same and s^i is type-dependent, the above theorem does not hold in any generality. Consider the following example:

Lemma 2 *Consider the two-period model with s^1 type-dependent and s^2 type-independent. If s^1 is observed when s^2 has to be observed, it is also observed when s^2 doesn't have to be observed.*

Proof. When the agent has to observe s^2, the value of observing s^1 decreases weakly both in the instrumental and ego-utility senses. □

When there is an unavoidable 'reality check' next period, we tend to hesitate more in making up our opinions today–being more afraid that our inferences will turn out wrong. Therefore, it seems that environments with good feedback can help make sure agents psychologically committed to believing in a judgment look at objective information carefully when reviewing that judgment, but promising that very feedback can undermine their willingness to make an assessment in the first place. Feedback provides a good learning environment for mistakes, but it is an environment that we might not want to enter.

3 A Psychological Sunk Cost Fallacy

As section 2 demonstrated, information-aversion (the tendency of agents to avoid information about their ability) has a variety of manifestations in information acquisition about

and the choice between different financial options. The opposite phenomenon, information-lovingness, would perhaps be less interesting in this setting, even if it wasn't the less common phenomenon anyway (Kőszegi 2000): since in that case both ego and instrumental considerations would compel the agent to accept free information, there won't be any (apparent) distortions in behavior. Thus it is not worth solving the model of section 2 for a convex u. Instead, to show one interesting phenomenon involving information-lovingness, I inject a little bit of my companion paper (Kőszegi 2000) into the previous model. Specifically, I first assume that ego utility takes a step-function form: above a certain cutoff, the agent is 'satisfied' with herself, and below it she isn't. Also, she has a 'threatened psychological stake' in one of the options: she made a subjective judgment in favor of this option, invested in it, and suffered a loss. She now has unsatisfactory beliefs about herself, and must decide whether or not to go on with the same option. I assume that the distribution of excess returns of her previous choice (relative to the other alternative) is centered around zero, and the agent can observe three identically distributed type-independent signals about it [9]. We keep the assumption from section 2 that financial payoffs are realized after ego utility. The characteristic financial behavior that results is a form of an apparent *sunk cost fallacy*: since the investor is psychologically committed to believing she has made the right call, she will tend to hold on to her losers too much. In the aggregate, those who change their choice do better financially than those who hold out with their old one. From this pattern, one might otherwise be tempted to conclude that the agents are acting irrationally, but here there is no real sunk cost fallacy going on–every participant does the financially sensible thing based

[9] These are just simplifying assumptions that allow us to abstract away from price determination specific to individual applications. For the result that agents hold losers too often, all we need is that news about the investment that would put beliefs on the 'good' side again are better than news that would make the agent hold. Assuming that without any new information the agent would want to hold, this is always the case. Theorem 5 wouldn't be meaningful without the symmetry assumption. And I'm using three signals so that after one signal, there could still be a reversal in which option the agent chooses.

on her information!

It's easy to see why this happens. By receiving a positive signal about the currently held option's relative return, the agent's past judgment will be put in a better light again, and this might put her back up on the positive side of ego utility. Thus, it is possible that she will stop acquiring information when she has seen a positive signal. In contrast, she will not stop acquiring information if she has seen a negative signal, since she has nothing to lose on the ego utility side. In addition, as in section 2, all agents do the financially sensible thing conditional on what they know: they choose the option with the higher expected return. This is the consequence of our assumption that financial outcomes are realized after ego utility. Therefore, those who end up selling trade on better information, and so on average will do better.

Theorem 5 *Consider an agent with negative beliefs about herself, holding an option whose excess return relative to the other option is distributed symmetrically around zero, and whose return is positively correlated with type. If the agent can observe three i.i.d. type-independent signals about the option's relative return, the expected relative performance of her choice conditional on selling is greater than the expected relative performance of her choice conditional on holding.*

Proof. If the agent doesn't stop acquiring information after one signal, the two expectations are clearly the same. If she does, those people who hold all have one signal to support this choice, while those who sell have one or three. Therefore, the expected profit of those who sell is greater. □

Theoretically, it is also possible that the agent *sells* a loser too easily. This can happen if her mean beliefs haven't dropped below the cutoff that makes her satisfied, and the beliefs that would make her hold are higher than the beliefs that would keep her satisfied about her ability. However, with mean beliefs above zero, the distortion could go either way: one could get too much or too little selling of losers.

As a final note on the last two sections, it is worth remarking that, somewhat surprisingly, information-aversion and information-lovingness often lead to similar observed behaviors: agents hold on to their assets too long. For self-image protection, this happens because the agent is unwilling to consider new information, and with self-image enhancement, because she wants information that, along with validating her past judgment, also justifies holding on to her investments.

4 Comment: Career Concerns or Ego Utility?

Although the agent's preferences in the models of this paper are non-standard, some of the resulting behavior can be explained in more traditional settings as well. Specifically, career concerns models can potentially give the agent similar incentives to those encountered here. For example, the fact that the agent fails to consider new information after having made a judgment in favor of an option is very reasonable for a manager with career concerns as well: throwing a bad light on previous judgments decreases the market's estimates of her ability.

In these situations, it is very hard to distinguish career concerns from 'ego concerns,' since the market payoffs can be set up to exactly match the ego utility function. However, there is at least one important way in which career-driven and ego-driven individuals differ: while a decisionmaker with standard preferences will always weakly prefer to receive *private* information, this is not necessarily the case when the ego enters the utility function as well. In other words, a career-driven manager will always want to get as much information as possible, she just might not want to reveal that information (or even the fact that she has received the information) to others. And my sense is that people are often reluctant to receive private information as well.

In addition, there are many situations in which career concerns explanations of the phenomena proposed in this paper are unreasonable. Such is the case with a small investor in the stock market, for example. It is very hard to imagine how one's review of one's choice

of stocks could reveal information to others that would affect the investor's success in any market.

5 Applications

5.1 Stock Market Participation

The model of section 2 applies readily to investors participating in the stock market. When choosing to invest in specific stocks, one has to make a variety of at least partially subjective judgments. Such judgments include not only how certain stocks can be expected to do based on the available information, but also what sorts of investments fit one's needs and which broker to trust for advice.

Since people are reluctant to base their decisions on subjective judgments, they might delegate such responsibility to a financial advisor, even if the financial advantages of such a decision are questionable. If investors do make a subjective judgment in favor of a stock and invest in it, they will try to avoid making further judgments and receiving more information on it. This is sluggishness. And even if an investor voluntarily or involuntarily finds that she might have made a bad decision, due to her psychological sunk cost she will try to convince herself that her decision wasn't so bad after all.

There is some evidence for each of these predictions of the model. Apparently, financial advisors offer little more than assurance and some hand-holding services: people who use them don't do better than those who don't. Recent work by Odean (1998) also indicates that small investors hold on to their losers too often: losing stocks people hold are drastically outperformed by the winners they sell, loosely in line with the model of section 3.

5.2 Small Businesses

The model of section 2 applies best to those already owning a small business. Due to sluggishness, they will try to avoid new information on their decisions as long as they can, leading them to respond too slowly to changing economic conditions. Theoretically, it is also possible that after having made a judgment about the business, the agent gets out too early to avoid receiving further signals about the quality of her judgment. However, this is only the case when getting out is less informative about previous judgments than staying in. The tasks associated with getting out, of course, also convey a lot of information about the quality of previous judgments; and, at the very least, this information could be delayed if one goes along casually managing the enterprise. Thus, although the agent might not *enter* into a business activity because of the self-image protection motive, this does not mean that she will *exit* because of it. What changes fundamentally when one enters is that a judgment is made about what to do, breaking the ex ante relation between information aversion and being out.

In addition, if the business does badly, according to section 3, agents will too often convince themselves that it's not so bad, staying in a losing enterprise for too long. Theorem 5 in section 3 also predicts that under some conditions, those who close up shop after suffering losses do better financially than those who continue, a result of the psychological commitment to earlier judgments about the business.

5.3 Project Choice by Managers

I have argued in my companion paper (Kőszegi 2000) that an empire-building ambition, understood as ego utility from the belief that one can manage big enterprises, can result in distorted project choices by managers. But whether or not project choice itself is distorted by the manager's ego utility, this should distort *how* she manages the chosen project. In particular, she is likely to be prone to all the distortions in instrumental decisions outlined

in section 2. Managers might want to hold off on decisions until it's clearer what to do (procrastination), but once they have made an important judgment call, they'll be reluctant to reexamine it.

If owners don't want their managers to fall prey to the above weaknesses, what can they do [10]? If the most important thing is to avoid procrastination, the results in section 2 suggest that hiring confident (even overconfident) managers might be a good idea. These people come in with enough conviction to affect the quick changes that might be necessary in the marketplace. A clear downside is that the changes could be excessive, at least if the manager is overconfident. To make things worse, confidence exacerbates the sluggishness problem, so the drastic changes could stick around for a long time [11]. Of course, there are things not modeled in the present paper that the owners might do to alleviate this problem; for example, they might want to switch management often, bringing in 'fresh blood' in the form of people whose egos are not threatened by a review of earlier judgments.

5.4 Extrinsic versus Intrinsic Motivation of Employees

Pay-for-performance systems necessarily involve distinguishing workers based on their performance, and there is a well-articulated notion that this might make those with a smaller bonus feel bad and eventually erode employee morale. In a model with ego utility, this has a natural meaning: the amount of compensation a worker gets is not only a signal about her performance, but indirectly also about her ability. Workers with a self-image protection motive don't like this and so have to be compensated for the 'ego pain' inflicted on them by the system. In addition, they want to take actions that make the bonus less informative about ability.

[10] Standard performance-based financial incentives are discussed in section 5.4.

[11] Fortune magazine's June 21, 1999 lead story on 'Why CEOs Fail' essentially identifies sluggishness as the number one CEO killer. Denial, as the magazine calls it, seems to be worst for very subjective decisions: the company's business model or subordinates selected for key positions.

This section examines one trick employers might do if they still want to provide incentives: they might purposefully use noisy signals to condition pay on, thereby drowning the inferences the employee can make about herself from her pay.

Formally, we build on the model of section 2. The agent can observe a type-dependent signal about project payoffs, with higher types getting more accurate signals. However, we make a few changes to make the point more easily. We assume there is only one period, but the amount of final compensation is revealed before the agent's ego utility is realized, so that it also provides information about ability [12]. To avoid the trivial solution that no incentives are necessary, we also assume that there is a utility cost $\epsilon > 0$ of observing the signal. And finally, the agent is now risk-averse: her utility function for monetary outcomes is v, where v is concave. Other than these changes, we operate in the neutral, independent setup of section 2.3.

This model applies more to certain kinds of jobs than to others. For example, in low or medium-level consulting jobs, it's not always clear what an individual's contribution to output is, and bonuses might well serve as the most reliable information on it. On the other hand, CEOs probably observe their own performance quite well, and so do workers in simple jobs where output has only a few dimensions, as is often the case in manufacturing [13]. In addition, there is little decisionmaking involved in the latter case.

We have to define the employer's problem. I assume that the employer observes the outcome a_i if the agent chooses project i, and can condition pay on this outcome. Let the conditional wages be w_0, w_1, and w_{a_1}. However, the owners of the firm can introduce some noise into the amount of compensation; in particular, if the agent chooses option 2,

[12] This makes the updating problem very similar to the two-period model, because there are two signals altogether (the type-dependent followed by the amount of compensation), which is informative about ability by lemma 1.

[13] This feature provides a testable prediction of the model, based on the incentive structure to be discussed below: naturally, incentives will vary depending on how applicable the model is.

they can mix the actual outcome with a purely random signal, paying w_1 with probability $p_H = tq_H + (1-t)q_L$ and w_0 with probability $1 - p_H$ [14]. Though formally equivalent, the best real-world interpretation of this assumption is probably that employers can choose not to remove noise from existing performance measures even if they are able to do so. I denote by α the probability that the actual outcome is used for pay; this is a choice variable of the employer. In setting up the problem this way, I'm ignoring the possibility that the employer 'cheats' in its assignment of bonuses, that is, avoids paying high wages somehow. Such a problem might be especially acute in this setting, where employees *don't want to* know how well they are doing. However, the employer's reputation and perhaps outside legal constraints can serve to restrain opportunistic behavior in assigning bonuses–the total wage bill is a good signal whether the employer followed the promised pay policy.

We assume that the employer wants to give the agent incentives to observe the type-dependent signal that is available to her, and choose option 2 if and only if that signal is good. Thus, we don't study the employer's full problem, only the implementation of one kind of agent behavior.

We start with the agent's updating problem. For $w_0 \neq w_1$, let

$$\begin{aligned} d_H(\alpha) &= Prob(q = q_H | s_1 = 1, wage = w_1) = \frac{cq_H\alpha + c(1-\alpha)p_H}{p_H} \\ d_L(\alpha) &= Prob(q = q_H | s_1 = 1, wage = w_0) = \frac{c(1-q_H)\alpha + c(1-\alpha)(1-p_H)}{1-p_H} \end{aligned} \tag{7}$$

$d_H(\alpha)$ and $d_L(\alpha)$ are the agent's posteriors about her ability when she observes a positive signal about project 2, chooses it, and then receives wages of w_1 and w_0, respectively. For example, when $\alpha = 0$ (when the payoff is based on a purely random signal,) $d_H(\alpha) = d_L(\alpha) = c$–the agent's payoffs are not informative about her ability. At the same time, $d_H(\alpha)$ is increasing and $d_L(\alpha)$ is decreasing in α, so that expected ego utility is decreasing

[14] This specific p_H is chosen for notational simplicity. It is the conditional probability of $a_2 = 1$ when receiving a good signal.

in α. Similarly, let $c_H(\alpha)$ and $c_L(\alpha)$ be the corresponding expressions when $s_1 = 0$. Note that, interestingly enough, $c_H(\alpha) < d_L(\alpha) < d_H(\alpha) < c_L(\alpha)$.

Clearly $w_0 \neq w_1$, otherwise the agent can't possibly have an incentive to observe the signal–she just chooses the option which leads to a higher (certain) payoff. Then the employer's problem is

$$\min_{\alpha} \frac{1}{2}w_{a_1} + \frac{1}{2}(p_H w_1 + (1-p_H)w_0) \text{ subject to}$$

$$\frac{1}{2}u(c) + \frac{1}{2}v(w_{a_1}) + \frac{1}{2}(p_H v(w_1) + (1-p_H)v(w_0)) + \frac{1}{2}(p_H u(d_H(\alpha)) + (1-p_H)u(d_L(\alpha))) - \epsilon \geq \overline{u} \text{ (IR)}$$

$$\frac{1}{2}u(c) + \frac{1}{2}v(w_{a_1}) + \frac{1}{2}(p_H v(w_1) + (1-p_H)v(w_0)) + \frac{1}{2}(p_H u(d_H(\alpha)) + (1-p_H)u(d_L(\alpha))) - \epsilon \geq$$

$$\max\left[v(w_{a_1}), \alpha\left(\frac{1}{2}v(w_1) + \frac{1}{2}v(w_0)\right) + (1-\alpha)(p_H v(w_1) + (1-p_H)v(w_0))\right] + u(t) \quad \text{(IC1)}$$

$$p_H v(w_1) + (1-p_H)v(w_0) + p_H u(d_H(\alpha)) + (1-p_H)u(d_L(\alpha)) \geq v(w_{a_1}) + u(c) \quad \text{(IC2)}$$

$$p_L v(w_1) + (1-p_L)v(w_0)) + p_L u(c_H(\alpha)) + (1-p_L)u(c_L(\alpha)) \leq v(w_{a_1}) + u(c) \quad \text{(IC3)} \qquad (8)$$

The principal wants to minimize the expected wages to be paid out subject to four constraints. The first of these is a standard individual rationality or participation constraint. The other three are incentive compatibility constraints, making sure that the agent observes the private signal about project payoffs. IC1 means that *before* observing the signal, the agent wants to see it and condition on it rather than relying on her previous information. At the same time, IC2 and IC3 amount to saying that *after* observing signals $s_1 = 1$ and $s_1 = 0$, the agent wants to choose options 2 and 1, respectively.

We solve the problem in several steps.

1. IC1 $\Rightarrow$ IC2. (easy)

2. We ignore IC3 and will see later that it is implied by the other constraints.

3. From IC1, $\alpha > 0$ and $w_1 > w_0$.

Intuitively, the posterior probability of the high outcome is higher after a good signal, so, in order to encourage the agent to choose option 2 in that case, the principal has to reward $a_2 = 1$ more, just as we would expect. This is easy to formalize.

4. IR binds-otherwise the employer can just decrease all rewards, still satisfying IC1.

5. IC1 binds-otherwise the employer can offer more insurance between w_1 and w_0, slackening the IR constraint.

6. $v(w_{a_1}) = \alpha\left(\frac{1}{2}v(w_1) + \frac{1}{2}v(w_0)\right) + (1-\alpha)(p_H v(w_1) + (1-p_H)v(w_0))$

   Proof. We simply prove that neither $v(w_{a_1}) > \alpha\left(\frac{1}{2}v(w_1) + \frac{1}{2}v(w_0)\right)+(1-\alpha)(p_H v(w_1)+ (1-p_H)v(w_0))$ nor $v(w_{a_1}) < \alpha\left(\frac{1}{2}v(w_1) + \frac{1}{2}v(w_0)\right) + (1-\alpha)(p_H v(w_1) + (1-p_H)v(w_0))$ is possible in an optimal solution.

   If the first one was the case, then one could offer more insurance between the outcomes w_1 and w_0 in a revenue neutral way, increasing the left-hand, but not the right-hand side of IC1.

   If the second was the case, then, since $w_1 > w_0$, we must have $w_1 > w_{a_1}$. Then decreasing w_1 and increasing w_{a_1} in a revenue-neutral way slackens both IC1 and IR. □

7. $v(w_{a_1}) = \overline{u} - u(c)$.

   Proof. Subtract IR from IC1, which both hold with equality. □

   This allows us to ignore w_{a_1} in the principal's minimization problem.

8. By point 6, IC2 holds strictly, the difference between the two sides being 2ϵ.

9. Without loss of generality assume that $\overline{u} = u(c) = 0$–we can do this by adding a constant to both v and u, if necessary. Then, from IR and IC1, respectively,

$$p_H v(w_1) + (1-p_H)v(w_0) \quad = \quad \Delta(\alpha)$$

$$\alpha\left(p_H - \frac{1}{2}\right)(v(w_1) - v(w_0)) \quad = \quad \Delta(\alpha), \tag{9}$$

where $\Delta(\alpha) = 2\epsilon - p_H u(d_H(\alpha)) - (1 - p_H)u(d_L(\alpha))$.

10. IC3 doesn't bind.

Proof. We prove that the left-hand side of IC3 is smaller than the right-hand side of IC2 by more than 2ϵ. This, together with point 8, implies our statement. The difference between the expected instrumental utilities is

$$(p_H - p_L)(v(w_1) - v(w_0)). \tag{10}$$

Using that $p_L = \alpha(1 - p_H) + (1 - \alpha)p_H$, this reduces to

$$\alpha(p_H - (1 - p_H))(v(w_1) - v(w_0)). \tag{11}$$

Now from above

$$\alpha\left(p_H - \frac{1}{2}\right)(v(w_1) - v(w_0)) = \Delta(\alpha). \tag{12}$$

Noting that $1 - p_H < \frac{1}{2}$, we have

$$p_L v(w_1) + (1-p_L)v(w_0) < p_H v(w_1) + (1-p_H)v(w_0) + p_H u(d_H(\alpha)) + (1-p_H)u(d_L(\alpha)) - 2\epsilon. \tag{13}$$

Finally, expected ego utility on the left-hand side of IC3 is negative, completing the proof. □

11. Solving the system 9 for $v(w_0)$ and $v(w_1)$ we get

$$\begin{aligned} v(w_0) &= \Delta(\alpha) - \frac{p_H}{\alpha\left(p_H - \frac{1}{2}\right)}\Delta(\alpha) \\ v(w_1) &= \Delta(\alpha) + \frac{1 - p_H}{\alpha\left(p_H - \frac{1}{2}\right)}\Delta(\alpha) \end{aligned} \tag{14}$$

The principal is interested in minimizing $p_H w_1(\alpha) + (1-p_H)w_0(\alpha)$, where we now take the wages to be functions of α. We can differentiate the above expressions for $w_0(\alpha)$ and $w_1(\alpha)$ and get

$$\begin{aligned} & p_H w_1'(\alpha) + (1-p_H)w_0'(\alpha) \\ = & \frac{p_H \Delta'(\alpha) - \frac{(1-p_H)p_H}{\alpha^2\left(p_H-\frac{1}{2}\right)}\Delta(\alpha) + \frac{(1-p_H)p_H}{\alpha\left(p_H-\frac{1}{2}\right)}\Delta'(\alpha)}{v'(w_1(\alpha))} \\ + & \frac{(1-p_H)\Delta'(\alpha) + \frac{(1-p_H)p_H}{\alpha^2\left(p_H-\frac{1}{2}\right)}\Delta(\alpha) - \frac{(1-p_H)p_H}{\alpha\left(p_H-\frac{1}{2}\right)}\Delta'(\alpha)}{v'(w_0(\alpha))} \end{aligned} \tag{15}$$

Let us start by examining the above derivative for two extreme cases. First, assume that the agent is (financially) risk-neutral, i.e. that v is linear. Then without loss of generality $v'(w_0(\alpha)) = v'(w_1(\alpha)) = 1$, so the derivative reduces to $\Delta'(\alpha)$, which is positive for any positive α [15]. This means that the employer wants to make α as small as possible, conditioning compensation a lot on a noisy signal of performance [16]. Although this is a highly unconventional result, in the context of this model it makes sense–since the agent doesn't care about financial risk but is averse to any real information on performance, the employer wants to drown the signal the incentives are based on in a lot of noise. With a risk-neutral agent, the principal can drown the signal in an arbitrarily large amount of noise, and by conditioning pay greatly on that noisy signal, still provide the incentives necessary.

At the other extreme, when the agent doesn't care about her ego or is 'information-neutral' ($\Delta'(\alpha) = 0$), and is also strictly risk-averse, the derivative 15 reduces to

$$\frac{(1-p_H)p_H}{\alpha^2\left(p_H-\frac{1}{2}\right)}\Delta(\alpha)\left[\frac{1}{v'(w_0(\alpha))} - \frac{1}{v'(w_1(\alpha))}\right] < 0 \tag{16}$$

since $w_0(\alpha) < w_1(\alpha)$. Consequently, in the optimal program $\alpha = 1$–when the agent doesn't care about her ego, we are back to the usual principal-agent problem, where adding noise

[15] It is easy to see that $\Delta(0) = 2\epsilon > 0$, $\Delta'(0) = 0$, and $\Delta''(\alpha) > 0$.

[16] For our purposes, it is not really important that for a risk-neutral agent the principal's maximization problem has no solution.

to the compensation is suboptimal. This would also be the case when the agent knows her type accurately. Indeed, piece rates are common in industries where the task is so simple it is unlikely workers would attach great personal importance to doing them well, and even if they do, they can't kid themselves for very long.

For an interior optimum, and assuming a well-behaved problem, the optimal α (α^*) is the solution to the equation

$$\begin{aligned} 0 &= \left[\frac{p_H}{v'(w_1(\alpha))} + \frac{1-p_H}{v'(w_0(\alpha))}\right]\Delta'(\alpha) \\ &+ \left(\frac{1}{v'(w_1(\alpha))} - \frac{1}{v'(w_0(\alpha))}\right)\left[\frac{(1-p_H)p_H}{\alpha\left(p_H-\frac{1}{2}\right)}\Delta'(\alpha) - \frac{(1-p_H)p_H}{\alpha^2\left(p_H-\frac{1}{2}\right)}\Delta(\alpha)\right] \end{aligned} \tag{17}$$

This equation summarizes the basic tradeoffs of the principal. If she increases α, the principal has to pay more in expected monetary utility to the agent because the agent's expected ego utility is lower. In other words, the principal has to compensate the agent for the extra information the payoff structure forces on her. This is the first term and tends to decrease α^*. The second two terms are related to the costliness of giving a risk-averse agent more incentives–to condition utility more strongly on the outcome while keeping expected utility the same, the principal needs to increase expected wages. The second term is the result of the fact that as α increases, it is more 'painful' for the agent to *follow* (as opposed to look at) her signal, so the principal has to give her more incentives to do it. This effect tends to decrease α^*. On the other hand, a higher α in itself provides better incentives, since pay is more a function of actual performance. This is represented in the third term, and tends to increase α^*.

Therefore, the optimal α balances the risk- and information-aversion of the agent. Consistent with this view, it is natural to conjecture that (holding u constant) if v is sufficiently risk-averse, then α^* is close to 1, and (holding v constant) if u is sufficiently information-averse, α^* is close to zero [17]. The following theorem implies both that the first of these

[17] It makes more sense to present this conjecture in a limit rather than a monotone comparative static

statements is false and the second one is true.

Theorem 6

$$\alpha^* \leq argmin_\alpha \frac{\Delta(\alpha)}{\alpha}. \tag{18}$$

Proof. The system of equations 9 that determines $w_0(\alpha)$ and $w_1(\alpha)$ is of the form

$$\begin{aligned} p_H v(w_1(a,b)) + (1-p_H)v(w_0(a,b)) &= a \\ v(w_1(a,b)) - v(w_0(a,b)) &= b. \end{aligned} \tag{19}$$

It is easy to see that $p_H w_1(a,b) + (1-p_H)w_0(a,b)$ is strictly increasing in a and, since v is concave, increasing in b. In the actual system 9, $a = \Delta(\alpha)$ and $b = \frac{\Delta(\alpha)}{\alpha}$. Since $\Delta(\alpha)$ is strictly increasing, for any $\alpha' > \text{argmin}_\alpha \frac{\Delta(\alpha)}{\alpha}$ the principal's expected payment is greater than for $\text{argmin}_\alpha \frac{\Delta(\alpha)}{\alpha}$. □

This theorem implies that a very information-averse agent will get very noisy incentives, irrespective of her risk-aversion. So, in some sense, the information aversion of the agent is a more important determinant of her incentive structure than her risk aversion is. What drives this result? Risk aversion makes using noisy signals very expensive, which should make reducing noise more important relative to protecting the agent's ego utility. However, in this problem compensating the agent for her loss in ego utility is not sufficient; if this is what the employer did, the decisionmaker would not observe her signal and just choose option 2. In order for her to choose option 2 if and only if her signal is good, she has to be rewarded more for the outcome $a_2 = 1$ relative to $a_2 = 0$, and she has to be rewarded more if the incentive structure is less noisy. This is also expensive to do for a risk-averse agent, and for a sufficiently information-averse agent the latter effect outweighs the former.

way. (That is, to say instead that α^* increases as v becomes more risk-averse, etc.) Comparative statics statements are will not be true in any generality because they depend on the third derivative of v.

Theorem 6, as well as the above discussion, applies only to 'judgment-sensitive' jobs, in which the agent has some private information (in this case her subjective judgment) which only she can use to take an action whose outcome depends on her ability. Even if the performance bonus reveals information about the agent's ability, but the job to be performed does not entail making a subjective judgment, the agent's financial risk aversion becomes relatively more important: in that case, it is sufficient to compensate the agent for the loss in ego utility inflicted on her, but there is no need to induce her to follow any private signal. This is an important distinction that implies that ego utility is likely to cause most problems in giving incentives to employees who have to collect and digest information and take actions based on it. This might be the reason why in industries in which people are paid to make judgments, employers are generally much more worried about their employees' ego.

The limit theorem 6 sets on the informativeness of incentives depends both on the information aversion of the agent and the disutility of the task. It is easy to prove that as ϵ approaches zero, $\text{argmin}_\alpha \frac{\Delta(\alpha)}{\alpha} \to 0$, so for the easiest or most enjoyable tasks, the incentives are very noisy, no matter how risk-averse the agent. Conversely, for a large ϵ, α will be close to one.

6 Conclusion

Building on the foundations of the psychology literature and arguments in my companion paper (Kőszegi 2000) in favor of a model based on ego utility, this paper considers implications of a concern for self-image for behavior. If agents acquire information about their ability indirectly through making judgments and observing their quality, and they have a self-image protection motive, they will be reluctant to observe combinations of signals that involve subjective judgments. Depending on whether a later or earlier signal is more valuable, this can lead them to sluggishness or procrastination, that is, not responding to new information or delaying making a subjective judgment, respectively. It can also lead to a

refusal to make decisions based on subjective judgments in the first place, relying instead on inferior objective information. In presenting applications to project choice by managers and intrinsic motivation, I also discuss how employers might try to alleviate the problems caused by self-image protection.

The next step is to explore how decisionmakers with self-image utility would fare in markets discussed in section 5 when there are other kinds of participants present as well. Taken literally, my model implies that ego-driven agents would be eliminated, because they don't use available information to make decisions. However, in a broader context this traditional objection clearly does not apply: the ego could affect other decisions than the information gathering one. For example, if ego utility makes agents confident (Kőszegi 2000), and confidence is complementary to effort, it might well be the agents without ego utility who are eliminated–agents with an ego work harder, and drive others out of the market. Once we realize that there is no strong reason to expect agents with ego utility to be out of markets, it would be interesting to analyze how market equilibrium is affected by their behavior.

A Proofs

Lemma 3 *Let current beliefs satisfy*

$$\begin{aligned} p_{0H} &= p_{1H} + \epsilon_H \\ p_{0L} &= p_{1L} + \epsilon_L \\ c &= Prob(q^t = q_H^t) = p_{1H} + p_{0H}. \end{aligned} \tag{20}$$

Then, a type-dependent signal about option 1 is uninformative about ability if and only if

$$(1-c)\epsilon_H(q_H^t - \frac{1}{2}) = c\epsilon_L(q_L^t - \frac{1}{2}). \tag{21}$$

Proof. By the law of iterated expectations, a necessary and sufficient condition for the beliefs about q^t not to move is to not have them move after the signal $s_1 = 1$. We have

$$\begin{aligned} Prob(q^t = q_H^t | s^1 = 1) &= \frac{p_{1H} q_H^t + p_{0H}(1 - q_H^t)}{p_{1H} q_H^t + p_{0H}(1 - q_H^t) + p_{1L} q_L^t + p_{0L}(1 - q_L^t)} \\ &= \frac{p_{1H} + \epsilon_H (1 - q_H^t)}{p_{1H} + \epsilon_H (1 - q_H^t) + p_{1L} + \epsilon_l (1 - q_L^t)} \\ &= \frac{\frac{1}{2} c + \epsilon_H (1 - q_H^t - \frac{1}{2})}{\frac{1}{2} + \epsilon_H (1 - q_H^t - \frac{1}{2}) + \epsilon_L (1 - q_L^t - \frac{1}{2})} \end{aligned} \tag{22}$$

Setting this equal to c gives the result. □

For $\epsilon_H = \epsilon_L = 0$, the result should be clear: if a_2 is independent of q and $a_2 = 0$ is just as likely as $a_2 = 1$, then no signal is informative about ability–in essence, the signal is the 'first' piece of information about the choice, and can't confirm or disconfirm previously held beliefs. It is, however, somewhat surprising that even if a_2 is not independent of q (ϵ_H, ϵ_L non-zero), the signal might not be informative. To understand this, consider $c = \frac{1}{2}$ and $\epsilon_H > 0$. One might think that $s^1 = 1$ is bad news: since it is more likely that $a_2 = 0$, chances are the signal is wrong, or that the agent is of low type. But if ϵ_L is sufficiently large, one wouldn't expect $s^1 = 1$ from low types, either, so the signal doesn't tilt beliefs in the negative direction.

Lemma 4 *Using the notation of lemma 3, a type-independent signal is uninformative about ability if and only if*

$$(1 - c)\epsilon_H = c\epsilon_L. \tag{23}$$

Proof. Similar to that of lemma 3.

Lemma 1 *A type-dependent signal followed by any other signal is always informative about ability.*

Proof. If the first signal is informative, then so are the two of them together–beliefs after the two of them are just a mean-preserving spread of the beliefs after the first. Thus, it is sufficient to prove that if the first signal is not informative, then the second one is.

To prove this for two type-dependent signal, first note that the condition in lemma 3 can only hold for both signals if $\epsilon_H = \epsilon_L = 0$ or the ratio $\frac{q_H - \frac{1}{2}}{q_L - \frac{1}{2}}$ is the same across the time periods. So unless this is the case, one of the signals is already informative, and order doesn't matter.

To complete the proof, we prove if one of the above two conditions holds, the relationship given in equation 21 can't be preserved after updating. The ratio corresponding to the ratio of ϵ_H and ϵ_L after updating is

$$\frac{p_{0H}(1 - q_H^t) - p_{1H} q_H^t}{p_{0L}(1 - q_L^t) - p_{1L} q_L^t} = \frac{\frac{1}{2}\epsilon_H - (q_H^t - \frac{1}{2})c}{\frac{1}{2}\epsilon_L - (q_L^t - \frac{1}{2})(1 - c)}. \tag{24}$$

If $\epsilon_H = \epsilon_L = 0$, a trivial use of lemma 3 shows that the second signal is informative. In the other case, the above ratio should be equal to $\frac{\epsilon_H}{\epsilon_L}$. For this we would need to have

$$\frac{\epsilon_H}{\epsilon_L} = \frac{c}{1 - c} \frac{q_H^t - \frac{1}{2}}{q_L^t - \frac{1}{2}}, \tag{25}$$

which, by lemma 3, is not true if the first signal is uninformative.

Now to prove that a type-dependent followed by a type-independent signal is informative, notice that the conditions of lemmas 3 and 4 can only hold at the same time if $\epsilon_H = \epsilon_L = 0$. So unless this is the case, we are done–otherwise, once again, one of the signals is informative by itself. And even if $\epsilon_H = \epsilon_L = 0$, the above proof shows that after updating using the period 1 signal, the posteriors won't satisfy the conditions of lemma 4. □

The proof takes advantage of the fact that in order for the first signal to be uninformative, the absolute value of ϵ_H has to be smaller than the absolute value of ϵ_L. But conditional on being type H, a signal is more informative, so ϵ_H is going to be moved by more than ϵ_L. Thus the ratio can't be preserved. □

Corollary 1 *Suppose that $a_1 > \frac{1}{2}$ and that the first-period signal is type-dependent and the second one is type-independent. If ego utility is sufficiently important (w is sufficiently large), only one of the signals will be observed. It will be the second one if and only if*

$$2(cq_H^1 + (1-c)q_L^1 - a_1)_+ < (q_I - a_1)_+. \tag{26}$$

Proof. We know from theorem 1 that for a large enough w exactly one of the signals will be observed.

If only one of the signals is observed, it will be the one with greater instrumental value. A signal has instrumental value if it can reverse a decision; and its value is the probability of reversing a decision times the difference in conditional expected utilities. With $a_1 > \frac{1}{2}$, a decision can only be reversed when the signal is favorable. Now both kinds of signals will equal 1 with probability $\frac{1}{2}$. It is also easy to show that

$$\begin{aligned} Prob(a_2 = 1 | s^1 = 1) &= cq_H^1 + (1-c)q_L^1 \\ Prob(a_2 = 1 | s^2 = 1) &= q_I. \end{aligned} \tag{27}$$

By fact 2, the decision will be reversed after $s^1 = 1$ if $cq_H^1 + (1-c)q_L^1 > a_1$ and after $s^2 = 1$ if $q_I > a_1$. The respective differences in expectations are therefore $(cq_H^1 + (1-c)q_L^1 - a_1)_+$ and $(q_I - a_1)_+$. Furthermore, since the first period's signal is there to affect both periods' choices, it will be preferred if it is at least half as valuable as the second. These conditions are summarized in inequality 26. □

References

Benabou, R., and J. Tirole (1999a): "Self-confidence: Interpersonal Strategies," Mimeo.

——— (1999b): "Self-confidence: Intrapersonal Strategies," Mimeo.

Carillo, J. (1997): "Self-control, Moderate Consumption, and Craving," Mimeo.

Carillo, J., and T. Mariotti (1997): "Wishful Thinking and Strategic Ignorance," Mimeo.

Fiske, S. T., and S. E. Taylor (1991): *Social Cognition*. McGraw-Hill, 2nd edn.

Gervais, S., and T. Odean (1999): "Learning To Be Overconfident," Mimeo, UC Davis.

Kőszegi, B. (2000): "Ego Utility, Overconfidence, and Task Choice," Mimeo.

Odean, T. (1998): "Are Investors Reluctant to Realize Their Losses?," Mimeo, UC Davis.

Rabin, M., and J. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics*, 114(1), 37–82.

Weinberg, B. A. (1999): "A Model of Overconfidence," Mimeo, Ohio State University.