

# EMOTIONAL AGENCY\*

BOTOND KÓSZEGI

This paper models interactions between a party with anticipatory emotions and a party who responds strategically to those emotions, a situation that is common in many health, political, employment, and personal settings. An “agent” has information with both decision-making value and emotional implications for an uninformed “principal” whose utility she wants to maximize. If she cannot directly reveal her information, to increase the principal’s anticipatory utility she distorts instrumental decisions toward the action associated with good news. But because anticipatory utility derives from beliefs about instrumental outcomes, undistorted actions would yield higher *ex ante* total *and* anticipatory utility. If the agent can certifiably convey her information, she does so for good news, but unless this leads the principal to make a very costly mistake, to shelter his feelings she pretends to be uninformed when the news is bad.

## I. INTRODUCTION

A common interaction of economic interest is one in which a party (call *her* the agent) has private information relevant to a decision both she and another party (call *him* the principal) care about. A large literature has explored outcomes and welfare both when the principal makes the decision based on communication with the agent, and when the agent does so herself.<sup>1</sup>

Economists have, however, mostly ignored one central aspect of the above types of interactions: in addition to its implications for the optimal action, the information the agent conveys also influences the principal’s *emotions*. Moreover, in many economically important situations, agents are cognizant of and respond to the presence of these emotions. Doctors want to provide sound medical advice, but do not want to scare patients with their prognoses. Executives want to make sensible managerial decisions, but do not want to impair morale in the process. Governments want to prepare their country for upcoming challenges, but do not want to induce panic or anxiety in citizens. And military

\* Comments by David Ahn, George Akerlof, Robert Anderson, Edward Glaeser, Steven Goldman, Jonathan Gruber, Theodore Keeler, Wei Li, Máté Matolcsi, Markus Möbius, Benjamin Polak, Luca Rigotti, Tanya Rosenblat, Emmanuel Saez, Chris Shannon, and seminar participants at CEU, the Harvard-MIT Public Finance and Health Economics Workshops, Rand Corporation, Tilburg University, University of California at Berkeley, and WZB contributed greatly to this paper. I would especially like to thank Edward Miguel, Matthew Rabin, and two anonymous referees for wonderful suggestions, and Andrew Hayashi for excellent research assistance.

1. See, for instance, Crawford and Sobel [1982] and Dessein [2002].

leaders, advisors, parents, friends, and lovers all take into account a soldier's, student's, or important other's emotional reactions to their behavior.

In the above settings, informed agents are confronted with a key dilemma: whether to ensure that chosen actions reflect *accurate* news, or to give the principal emotionally beneficial *good* news. This paper studies agents' behavior in the face of such a dilemma.

Section II presents the model, which extends Caplin and Leahy [2004]—a theory of communication with emotionally relevant certifiable information and no choice of action—to a cheap-talk setting, and to a situation where the agent's information has decision-making value. To isolate the question of how market participants respond to emotions from (other) principal-agency issues, I assume that there is no conflict between the parties: the agent's goal is to maximize the principal's utility. The principal's utility is a convex combination of two parts: future utility from physical outcomes, and current anticipatory emotions, which depend on rationally formed beliefs about the *exact same* outcomes. Both components take the expected-utility form, and  $w$  is the weight on emotions in the principal's total utility.<sup>2</sup>

The two parties face the following situation. The agent privately observes a state of the world  $s$ , which affects the achievable level of physical utility as well as the physical-utility-maximizing action. After receiving a message from the agent, the principal chooses one of two actions  $t$ , and experiences anticipatory utility based on his current beliefs.<sup>3</sup> Later, I also consider a setting where the agent chooses  $t$ . Section III presents evidence and arguments that this model captures an important ingredient of the medical, political, managerial, and personal situations mentioned above. The rest of the paper analyzes the model, and uses these applications to illustrate the results.

In the majority of applications, the agent's information  $s$ —often based on subjective judgments or sensitive material—is impossible to certifiably communicate, so only cheap talk about it is

2. For presentational purposes, the paper will assume throughout that the agent cares directly about the principal's emotions and does so just as much as he does. As I note below, the results would remain qualitatively unchanged if the agent cared less about the principal's feelings, or even if she cared only indirectly about them due to their effect on physical outcomes.

3. Formally, the principal's and the agent's behavior, and the principal's beliefs, are determined in an "emotional perfect Bayesian equilibrium," where the standard concept is appropriately modified for games with emotions.

feasible. Subsection IV.A proves that in this case, essentially the only information the agent can communicate is the action she recommends. In addition, if a particular action tends to be physically optimal when the state is favorable, to increase anticipatory utility she distorts her recommendation toward this action. For example, to inculcate some hopeful feelings a doctor might recommend fighting an advanced cancer, even though physical health does not warrant the intervention. Knowing the agent's motivations, however, the principal realizes that recommendations are distorted from a physical-utility point of view, and because ultimately his emotions derive exactly from expected physical outcomes, his *ex ante* expected anticipatory utility is also not maximized. Due to emotions, therefore, the parties cannot achieve their joint goal of maximizing the principal's expected utility based on the agent's information. Moreover, strong feelings have a self-defeating aspect: the higher is the weight  $w$  on emotions, the stronger is the agent's incentive to manipulate the principal's beliefs, and therefore the lower is average physical *and* anticipatory utility.

Subsection IV.C considers further comparative statics. I show that when the achievable level of physical utility does not depend on  $s$ , when choosing the wrong action is very costly, and when the principal is ignorant about the decision-making problem, expected utility is actually at or close to the optimum. And subsection IV.D argues that welfare may be increased if the agent publicly commits to "norms" that specify what to do in each situation, or even if she commits to doing the same thing virtually independently of  $s$ . For example, the above logic indicates that following 9/11, the government would without some form of commitment give out reliable and hence fear-provoking warnings of an imminent terrorist attack way too rarely. Since a truly informative warning system cannot be maintained, the Department of Homeland Security makes a point of keeping the threat level high even when little information indicating danger is available.

In Section V, I assume that the agent can certifiably reveal  $s$  if she knows it, but still cannot certifiably show that she does *not* know  $s$ . This assumption applies to information that can be documented or verified, and to personal or other relationships in which the agent would for moral or legal reasons not tell an outright lie, but is not obligated to reveal everything she knows. I start by abstracting away from action choice, making my setup similar to Caplin and Leahy [2004]. When the agent observes  $s$

with probability 1, a generalization of their full-disclosure result holds: because “no news” would be interpreted by the principal as “bad news,” the agent reveals all of her information. When the probability that the agent is uninformed (does not know  $s$ ) is positive, however, I show that she reveals relatively favorable news she has, but not bad news she has. For example, if a manager has indications that her team’s project is not going so well, she might protect her employees’ morale by not telling them about this.

Finally, I reintroduce action choice into the above model. There may then be a new type of equilibrium, in which the agent discloses good news and not mediocre news, but to ensure the principal makes the right choice also discloses very bad news. To continue with the management example, if the project is going so poorly that its successful conclusion is in danger, the manager may be blunt with employees to make sure the problem is addressed.

While the models in this paper capture a basic dilemma regarding emotions and information in relationships, they leave open the question of what happens, for instance, when parties have multiple opportunities to communicate, or when the principal is naïve in interpreting the agent’s behavior. Section VI discusses this and other potential extensions of the current theory, and concludes.

## II. THE MODEL

This section introduces the model. Following the methodology of Caplin and Leahy [2001, 2004], I set up the principal’s utility as a function of physical outcomes and beliefs-based emotions. I adapt their framework to allow for cheap-talk communication, and for an action to be taken by one of the parties. In addition, to make the intuition behind results correspond more closely to the mathematical analysis, the model uses a different solution concept from Caplin and Leahy’s [2004].

### *II.A. The Utility Function*

I begin with a discussion of the principal’s utility function, which is intended to capture “instrumental” concerns usually stressed in economics, as well as anticipatory emotions that respond to information. There are two periods, 1 and 2, and total utility is a convex combination of actual and anticipated instrumental/physical outcomes in period 2. The principal’s period 2

physical utility takes the form  $h(s,t)$ , where  $s \in [0,1]$  is the state of the world realized in period 1, and  $t \in \{0,1\}$  is an action taken in the same period. The state  $s$  is continuously distributed with positive density everywhere.<sup>4</sup> Depending on the application, the action  $t$  may be taken by the principal, or by the agent on the principal's behalf. The basic model covers the former case, and subsection IV.B will identify how the latter one is different. The paper will consider several specifications for  $h(s,t)$ .

In the first period, the principal derives utility from the anticipation of period 2 physical outcomes. Specifically, I assume that his anticipatory utility is equal to his expected physical utility in period 2 conditional on his beliefs  $\mu$  in period 1, taking the form,

$$\int h(s',t) d\mu(s').$$

The beliefs  $\mu$  will be determined endogenously as part of the equilibrium.

The principal is an expected-utility maximizer with a von Neumann-Morgenstern utility function that is a weighted sum of physical and anticipatory utility:

$$(1) \quad U(\mu,s,t) = w \cdot \int h(s',t) d\mu(s') + (1-w) \cdot h(s,t),$$

where  $w \in (0,1)$ . While linearity of emotions in the principal's beliefs is not crucial for the qualitative results of the paper,<sup>5</sup> it is practical from a methodological point of view. It implies that for any  $w$  the principal maximizes expected physical utility in any individual decision-making problem, meaning both that the results are due to the *agent's* responses to the principal's emotions rather than his responses, and that  $w$  parameterizes the impor-

4. In this formulation, all uncertainty is captured in  $s$ , and since the agent will be assumed to observe  $s$ , she has full knowledge of the principal's future physical utility. More generally interpreted, the same framework also captures situations in which the agent has residual uncertainty regarding the state of the world as well. In that interpretation,  $s$  is a signal observed by the agent, and  $h(s,t)$  is the principal's expected physical utility conditional on  $s$ .

5. In particular, I have considered a model in which the principal's anticipatory utility can be a nonlinear function of expected physical utility. While the description of the ex ante optimal policy is more complicated in that case, the key result of subsection IV.A—that the action associated with good news is chosen too often relative to the ex ante optimal policy—still holds. All the qualitative results in later sections also survive.

tance of emotions for principals with “otherwise identical” preferences.

For conceptual clarity, the model treats anticipatory utility as completely separate from physical utility. Section III below notes that in reality emotions often affect relevant physical outcomes, and emphasizes that this in fact widens the model’s applicability because it implies that even an agent who does not care directly about the principal’s emotions may need to take them into account.

Most of this paper will assume that the attainable level of physical utility varies with  $s$ , and that the optimal action varies systematically with the attainable level of physical utility.

ASSUMPTION 1 (A1).  $h(s,t) = s - l(s,t)$ , where  $l(s,t) \geq 0$  for all  $s$ ,  $t$ , and  $\min_{t \in \{0,1\}} l(s,t) = 0$  for all  $s$ . Furthermore,  $L(s) \equiv l(s,1) - l(s,0)$  is strictly decreasing and continuous, and there is an  $s^* \in (0,1)$  such that  $L(s^*) = 0$ .

(A1) implies that for any state of the world  $s$ , the attainable level of physical utility is  $s$ . Relative to this optimum, the principal is—as determined by the “loss function”  $l(s,t)$ —made physically worse off if the wrong action is taken. In addition, the more favorable (higher) is the state of the world, the more appropriate is the action  $t = 1$  relative to  $t = 0$ . In the medical arena, for example, the following two very different situations are both consistent with (A1).

1. A patient’s blood sugar is measured. If it is in the healthy range ( $s$  is high), he can live as he used to ( $t = 1$ ). But if it is too high ( $s$  is low), a change in lifestyle ( $t = 0$ ) is warranted.
2. A prognosis is made for a patient known to have cancer. If there is a reasonable hope of survival ( $s$  is high), aggressive treatment ( $t = 1$ ) is called for. If he is in the terminal stage ( $s$  is low), the best option is palliative care ( $t = 0$ ).

Section III argues that the properties of  $h$  are also consistent with the other applications of emotional agency. But there are also (medical and other) situations that do not fit into this framework, so in subsection IV.C, I discuss how alternative specifications of  $h$  modify the results.

## II.B. *The Communication Game*

The previous section has formulated the principal’s preferences, including his utility from anticipatory emotions. The pri-

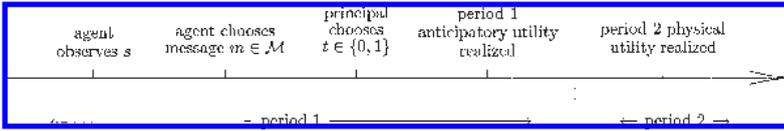


FIGURE I  
Timing

mary interest of this paper, however, is in studying how *other* economic decision-makers respond to the presence of anticipatory utility. I therefore introduce the agent into the picture. I assume that her interests are perfectly aligned with those of the principal, so that the von Neumann-Morgenstern utility function of both parties is given by equation (1). The paper's qualitative results would remain unchanged if the agent did not fully internalize the principal's emotions, but cared about them at least to a certain extent, or for whatever other reason preferred to convey better rather than worse news.

Given this and the earlier claims of greater generality, the paper's modeling strategy is worth highlighting. The goal in the specification of utility is not perfect realism or complete generality, but the isolation of the agent's response to emotions from issues not directly related to emotions. By assuming that the parties have the same utility function, the model abstracts from standard problems—explored in the principal-agent and communication literatures—that arise due to decision-makers' divergent private interests.<sup>6</sup> And by assuming that anticipation ultimately depends on expected utility from physical outcomes, the setup holds constant preferences over those standard outcomes.

The agent and the principal play the following game, whose timing is illustrated in Figure I. First,  $s$  is drawn and is observed by the agent, but not by the principal. One can make two extreme assumptions regarding the agent's ability to communicate  $s$ . The first and larger part of the analysis, in Section IV, assumes that she cannot certifiably reveal  $s$  itself, and the second part, in

6. Two strands of the standard information-economics literature are related to the models of this paper. First, the assumption that the agent needs to guide the principal in taking an action is a feature shared with models of expert advice [Crawford and Sobel 1982; Scharfstein and Stein 1990; Prendergast and Stole 1996; and others]. Second, disclosure decisions—which are investigated in the context of emotions in Section V—have also been studied in corporate finance [Grossman and Hart 1980; Milgrom 1981; Jung and Kwon 1988].

Section V, considers the opposite case. The former assumption is plausible when—as in most applications of emotional agency—the agent’s information is subjective or very difficult to quantify and prove. In this case, the game continues with the agent choosing a cheap-talk message  $m$  from a finite set of messages  $\mathcal{M}$ . Upon observing  $m$ , the principal chooses  $t$ , after which his anticipatory utility is realized based on his momentary beliefs. Finally, the principal’s period 2 physical utility is realized. Subsection IV.B analyzes the variant of the model in which the agent chooses  $t$ .

As a solution concept for models in this paper, I employ a version of perfect Bayesian equilibrium, *emotional perfect Bayesian equilibrium (EPBE)*. While the standard definition of perfect Bayesian equilibrium applies to games with instrumental utility only, its requirements are easily adapted for the present game. First, the principal chooses  $t$  optimally given his beliefs about  $s$ . Second, for any realized  $s$ , the agent chooses her message  $m$  to maximize the principal’s utility, correctly anticipating what inferences he will draw from  $m$ , and what action he will choose. Third, the principal forms his beliefs using Bayes’ rule to infer the meaning of the agent’s message, knowing her strategy. Since the formal definition of EPBE requires extra notation and is not necessary to understand the results, it is relegated to the Appendix.<sup>7</sup>

I will refer to an EPBE as “simple” if two messages that induce the same probabilistic action and the same anticipatory utility are in fact the same message. Notice that we can “transform” any EPBE into a simple one by using a single message in place of any such equivalent messages.<sup>8</sup> Since the analysis in this paper centers around the principal’s emotions and the distortions

7. In their paper, Caplin and Leahy [2004] use the “psychological equilibrium” solution concept of Geanakoplos, Pearce, and Stacchetti [1989]. In psychological games, players derive utility from their (possibly higher-order) beliefs about others’ strategies. Appropriately defined, psychological equilibrium and EPBE yield identical predictions in the model above, since the principal’s beliefs about the state of the world following different messages derive from his beliefs about the agent’s strategy. Since the principal’s utility depends directly on beliefs about the state of the world, however, the EPBE solution concept corresponds more closely to the intuition driving the results.

8. For any set of messages  $M \subset \mathcal{M}$  such that all messages in  $M$  induce the same behavior and the same anticipatory utility in the principal, select some arbitrary  $m \in M$ . Then, change the agent’s strategy so that whenever she would send a message in  $M$ , she sends  $m$ , adjust the principal’s beliefs accordingly, and leave everything else unchanged. It is easy to check that this still satisfies the requirements of an EPBE.

in action choice, this transformation changes nothing of interest. Hence, I restrict attention to simple EPBE.

### III. APPLICATIONS

This section presents possible applications of the model. For each application I argue that the model—though too simple for a full description—captures some of the interaction’s crucial aspects. The results in later sections will be applied to and illustrated by these examples.

#### III.A. *Doctor-Patient Interaction*

The primary role of a physician is to identify patients’ state of health and provide advice and help to prevent and treat medical problems. By their very nature, however, the advice and news provided by a doctor powerfully influence patients’ fears, hopes, and myriad other emotions aroused by medical situations. As a result, doctor-patient communication is a natural application for the model. In this application the agent is the doctor, and the principal is the patient. The state of the world,  $s$ , becomes a diagnosis observed by the doctor, and her message  $m$ , a recommendation about what treatment to choose or lifestyle to follow. Finally, the principal’s anticipatory utility corresponds to the patient’s emotions regarding his future health.

There is ample evidence that doctors are cognizant of patient emotions, and try to take them into account when providing care [Ptacek and Eberhardt 1996; Christakis 1999; Kőszegi 2004]. Importantly, physicians’ concern for patients’ feelings is not merely about emotional management—but rather a crucial ingredient in achieving the best *physical* health. Emotional states affect physical outcomes both directly through immunological mechanisms [Damasio 2000, p. 120], depression, and insomnia,<sup>9</sup> and indirectly through influencing compliance and other behavior, or even overwhelming the patient so much that she cannot make reasoned choices [Tulsky 2004].<sup>10</sup> Thus, even doctors who care exclusively about a patient’s medical condition are ultimately forced to take his feelings into consideration as well.

9. Specifically, for instance, Todaro et al. [2003] find that negative emotions increase the frequency of coronary heart disease.

10. See Kőszegi [2004] for further evidence on these issues.

### III.B. Political Economy

Governments and social planners typically have better information than most citizens about their country's upcoming challenges in the economic, social, and security spheres. One of a government's important jobs is to take measures based on this information to increase the welfare of the country. As an extreme example, the U. S. government collects information on the terrorist threats facing the nation, and—by means of its color-coded threat advisory system—can set expensive processes in motion to decrease the likelihood of an attack. The Federal Reserve makes similarly high-profile decisions regarding the economy, and numerous smaller and more localized warning systems (regarding storms, air quality, etc.) exist in many places.

There is little doubt that citizens observe and interpret these kinds of decisions, and draw conclusions that (at least for important issues) affect their anticipatory utility. Indeed, one of the major considerations in issuing a terrorist warning seems to be its judged effect on fear in the population.<sup>11</sup> As in the case of doctors, a government's interest in the population's feelings is about much more than emotional management. Most importantly, leaders know that fears of a terrorist attack or another disaster would interfere with people's work and everyday life, and could thus cripple the economy. For example, SARS was widely viewed as devastating to local economies in a large part due to the population's emotional reactions to the danger.<sup>12</sup>

This situation also fits the model. The government (agent) observes an economic or environmental forecast or intelligence information ( $s$ ). This information affects the optimal subsequent action ( $t$ ), and is relevant for the population's (principal's) antici-

11. The Department of Homeland Security itself recognizes that “[r]aising the threat condition has economic, physical, and psychological effects on the nation” (<http://www.dhs.gov/dhspublic/display?theme=29>). The same view is repeatedly echoed by journalistic writings on the terror alert system. Regarding a decision of whether to issue a warning in 2002, for instance, Jeanne Cummings and Gary Fields write that “officials struggled to strike the balance between warning Americans of a potential strike and avoiding panic-provoking false alarms” [*Pittsburgh Post-Gazette*, May 19, 2002, page A-12].

12. For example, prime minister Goh Chok Tong of Singapore expressed publicly that the fear generated by SARS could “kill” an economy. A Toronto Dominion Bank report estimated the cost of SARS to the Canadian economy at above GBP 1 billion [Tara Womersley, *The Scotsman*, April 30, 2003, page 3]. The fact that officials performed largely symbolic acts—such as dining in affected areas—to reassure the population indicates that a lot of the damage was due to fears. Underlining this view, the *Christian Science Monitor* writes: “The history of public health has shown repeatedly that the social and political responses to an outbreak of disease will diverge from its medical realities” [June 5, 2003, page 9].

patory emotions. In contrast to the doctor-patient application, however, it is often the government that decides on  $t$ , instead of recommending it to the population.

### *III.C. Morale in Organizations and in War*

Managers believe that if morale suffers, productivity suffers [Bewley 1999]. Although many factors likely affect morale, one of its main sources is an employee's perception of his ability and of his career prospects at the firm.<sup>13</sup> A manager (agent) often has a better understanding ( $s$ ) than an employee (principal) about his prospects, as well as the job-related activities ( $t$ ) he should undertake. This applies to wide-reaching decisions such as whether to prepare for a possible promotion, as well as everyday challenges such as which tasks to perform. The optimal action typically reflects on the employee's ability and prospects, affecting his morale and consequently also his productivity. Finding out, for instance, that one is not sufficiently capable to prepare an important presentation can be quite demoralizing. Depending on whether the manager gives orders regarding what to do or merely communicates her views, the situation fits either the version of the model in which the agent chooses  $t$ , or the version in which the principal chooses  $t$ .

Morale is also vital to success in military conflict. Without a firm resolve to support their country's struggle, few soldiers can be motivated to take the kinds of risks, and civilians to make the kinds of sacrifices, that an effective war effort involves. An important component of morale is the population's perception of how the war is progressing. If people believe matters are bogging down, they "subconsciously prepare for defeat," a tendency that is true of both civilians [Broughton 1942] and soldiers [Gurfein and Janowitz 1946]. A leader (agent) is better-informed ( $s$ ) than most soldiers and citizens (principal) regarding the details of the war, and bases many of her decisions ( $t$ ) on this information. In the process, she must take into account how the interpretation of her observed behavior will affect the country's morale. Thus, this

13. Miller [1941], Rose [1950], Bluedorn [1982], and Johnsrud and Rosser [1999] provide evidence for and discuss this aspect of morale, and Fang and Moscarini [2005] study its implications for wage compression when workers are overconfident about their ability. More comprehensively, morale is usually defined as employees' state of mind and sense of common purpose with respect to their work (Zeitz [1983], for instance).

example fits the variant of the model in which the agent chooses  $t$ .

### *III.D. Parenting, Personal Relationships, and Advising*

The model also applies to several more personal situations of tremendous importance for welfare.

A child's life, particularly in teenage years, is characterized by powerful hopes and anxieties about his ability and future. Due to a parent's dual responsibility of managing such emotions and at the same time guiding her children toward the right choices, she is likely to regularly encounter situations of the type in this paper. For instance, a child (principal) might have to decide how much effort ( $t$ ) to devote to an educational or athletic activity where his prospects invoke strong anticipatory emotions in him. A parent (agent) often has an opinion ( $s$ ) about the child's prospects, and these are systematically related to what he should do: if he is not good enough, he should put more of his energies elsewhere.

Consider also romantic attachment and friendship. A lover (agent) often has an opinion ( $s$ ) about her partner (principal) based on which the relationship could be improved. To take a common example, she may find him inattentive, and prefer that he take steps ( $t$ ) to improve on this front. Knowing about her dissatisfaction, however, may make him feel bad about himself and about the relationship. Both because she cares about her partner's well-being and because such emotions can have an effect on the relationship itself, she is likely to take his feelings into account.

Finally, the model also applies to academic advising. A student (principal) might show up at a professor's (agent) office with an idea, looking for advice ( $m$ ) on whether to pursue ( $t$ ) the idea. The professor forms an opinion ( $s$ ) about the quality of the proposed research, and should advise the student to keep working on it only if it is promising. Unfortunately, if the student believes that his ideas are mediocre, he may become depressed and unmotivated to do work.

## IV. RESULTS

### *IV.A. Equilibria When the Principal Chooses the Action*

This section analyzes the basic model, in which the agent can reveal information only through cheap-talk messages, and the

principal chooses  $t$ . This model applies to medical situations where the doctor's information is not merely a set of objective test results, to nonbinding government recommendations such as travel and health advisories, and to much subjective managerial, academic, and personal advice on how to lead one's life. The key implication of the model is that despite perfectly aligned interests, the agent's response to the principal's emotions leads her to recommend the action associated with bad news too rarely relative to the welfare-maximizing strategy.

The following lemma greatly simplifies the analysis.

**LEMMA 1.** Suppose that (A1) holds. In a simple EPBE, the agent sends at most two messages with positive probability. If she sends exactly two messages with positive probability, one of the messages leads the principal to choose  $t = 0$  with probability 1.

Lemma 1 says that the information the agent is able to convey is endogenously limited to at most a binary message. The intuition is in three parts. First, in a simple EPBE different messages induce the principal to choose  $t = 1$  with different probabilities, and so can be thought of as different "recommendations" that carry no information beyond the suggested action. If they did, the agent would always convey the best possible news along with a given recommendation. Second, because this maximizes both physical utility and anticipatory utility, for all  $s > s^*$  the agent gives the recommendation that maximizes the probability of  $t = 1$  being taken. Third, any other message therefore reveals that  $s \leq s^*$ , leading the principal to choose  $t = 0$  with probability 1. And by the first point above, in a simple EPBE there can only be one such other message.

As in other cheap-talk games, a "babbling" EPBE, in which  $m$  is uninformative of  $s$  and does not affect the principal's choice of  $t$ , clearly always exists. An EPBE in which the principal chooses  $t = 0$  after one message and randomizes after the other one is effectively also an uninformative EPBE, since it leads to the same expected utility as choosing  $t = 0$  regardless of  $s$ .<sup>14</sup> Hence, by Lemma 1, the only type of EPBE in which the agent's information helps the principal in choosing  $t$  is one where she sends two

14. To see this, note that whenever the principal randomizes, he is indifferent between  $t = 0$  and  $t = 1$ . Thus, his expected physical utility is the same if he always chooses  $t = 0$  instead of randomizing. And as argued below, total expected utility is equal to expected physical utility.

messages with positive probability, and these messages correspond to recommendations to take, respectively,  $t = 0$  or  $t = 1$  with probability 1.

I now investigate the agent's behavior in these "informative" EPBE when they exist, and return to the conditions for existence later. Suppose therefore that, as is necessary for an informative EPBE, the principal follows each of the agent's recommendations.

Notice that by the law of iterated expectations, the principal's ex ante expected anticipatory utility is equal to his expected physical utility, so that both of these components are also equal to his ex ante expected total utility. This implies that in all versions of the model the ex ante optimal policy is to always maximize physical utility, in the process also maximizing expected anticipatory utility.<sup>15</sup> In particular:

**PROPOSITION 1 (Ex Ante Optimal Policy).** Suppose that (A1) holds.

Then, the ex ante optimal policy is to commit to recommending  $t = 0$  if and only if  $s \leq s^*$ .<sup>16</sup>

With the agent free to choose her recommendation, however, she recommends  $t = 0$  too rarely relative to this optimal policy:

**PROPOSITION 2 (Distortion in Action).** Suppose that (A1) holds. In any informative EPBE, there is an  $s^c \in (0, s^*)$  such that  $t = 0$  is chosen if and only if  $s \leq s^c$ .

Proposition 2 means that in any EPBE, the agent maximizes neither expected physical utility nor (by implication) ex ante expected total utility—even though maximizing total utility is her goal! To gain intuition for this result, suppose that the principal believes that the agent is following the ex ante optimal policy, and she observes an  $s$  close to but below  $s^*$ . From a physical-utility point of view, she is then approximately indifferent as to what she should recommend. But in consideration of the principal's emotions, she strictly prefers to recommend  $t = 1$ , because it imparts the good news that  $s > s^*$  rather than the bad news that  $s \leq s^*$ . In EPBE, this motivation to increase anticipatory utility at the expense of physical

15. One convenient feature of the linearity of anticipatory utility in beliefs is that the optimal ex ante policy is so simple to describe. For other forms of anticipatory utility, it would not in general be true that the optimal policy maximizes expected physical utility. As noted in footnote 5, however, the key results in the paper would still hold in that case.

16. To avoid clumsy constructs relating to "if and only if" statements, I assume when indifferent the agent recommends or chooses  $t = 0$ , and in Section V discloses  $s$ .

utility leads the agent to expand the range of states for which she recommends  $t = 1$ . But because ultimately emotions are about expected physical utility, such behavior only decreases the principal's EPBE expected utility.<sup>17</sup>

As the above intuition indicates, both Proposition 2 and its implication for welfare are results of the agent's response to the principal's emotions. Indeed, if the principal's utility did not include an anticipatory component ( $w$  was equal to 0), there would be a perfect Bayesian equilibrium in which the agent recommends  $t = 0$  if and only if  $s \leq s^*$ , maximizing welfare.

In a certain sense, the agent cannot maximize the principal's expected utility not *despite*, but exactly *because of* trying to do so. Specifically, if she cared solely about the principal's physical utility, she would give him ex ante optimal recommendations from both the physical-utility and the anticipatory-utility points of view. But this is likely to be a rare case in practice: in the medical, political-economy, morale, and advising applications, where the agent may not care directly about the principal's emotions, she does care indirectly about those emotions because they influence pertinent physical outcomes.

The cutoff state  $s^c$  identified in Proposition 2 must satisfy the condition that the agent is indifferent between recommending  $t = 0$  and  $t = 1$  when the state is  $s^c$ . This is equivalent to

$$(2) \quad w \cdot (E[s - l(s, 1) | s > s^c] - E[s | s \leq s^c]) = (1 - w) \cdot L(s^c).$$

Equation (2) summarizes the key trade-off driving the agent's behavior, that between providing *good* news and providing *accurate* news. Knowing that the state is  $s^c < s^*$ , the recommendation that maximizes physical utility—the accurate news—is  $t = 0$ . But recommending  $t = 1$  makes the principal feel better—and hence is good news—because he infers that  $s > s^c$  instead of concluding that  $s \leq s^c$ . The right-hand side of equation (2) is the physical-utility gain of giving accurate news, while the left-hand side is the anticipatory-utility gain of giving good news. At the cutoff  $s^c$ , the two considerations must cancel.

The trade-off embodied in equation (2) also clarifies why the

17. The intuition for Proposition 2 is similar to the time inconsistency of government policy identified by Kydland and Prescott [1977]. Expectations regarding future policy influence economic actors' current decisions, so unless the government can commit to future policy choices, these choices do not maximize the current social welfare function. As the authors' analysis itself reveals (see page 476), however, if economic actors aim to maximize the social planner's objective function, no time inconsistency arises.

agent treats principals with different  $w$ 's—who behave identically in all individual decision-making problems—differently. From the point of view of the principal, the distribution of future outcomes and his beliefs about those outcomes always coincide. Since maximizing his anticipatory utility and maximizing his physical utility are therefore equivalent, the principal's behavior is independent of the weight he puts on each component, always maximizing expected physical utility conditional on his beliefs. From the point of view of an agent who knows the principal's future physical utility better than he does, the distribution of physical outcomes and the principal's beliefs about those outcomes do not coincide. Because there is therefore a trade-off between the anticipatory and physical components of utility, the agent's behavior depends on  $w$ .

In addition to equation (2), in an informative EPBE an incentive-compatibility constraint must be satisfied: the principal must be willing to follow each of the agent's recommendations. When the recommendation is  $t = 0$ , he is willing to do so, because in that case he infers that  $s < s^*$ . When the recommendation is  $t = 1$ , he is willing to follow it if and only if  $E[l(s,0)|s > s^c] \geq E[l(s,1)|s > s^c]$ . Based on these conditions, the following lemma characterizes informative EPBE.

LEMMA 2. Suppose that (A1) holds. There is an EPBE such that  $t = 0$  is chosen exactly when  $s \leq s^c$  if and only if  $s^c$  satisfies equation (2) and  $E[l(s,0) - l(s,1)|s > s^c] \geq 0$ .

While there are often multiple EPBE, since the parties have aligned interests it may be natural for them to coordinate on the best one. The following proposition analyzes welfare in these best EPBE as a function of the weight  $w$  on emotions, and considers the existence of informative EPBE.

PROPOSITION 3 (Welfare and Existence of Informative EPBE). If (A1) holds:

- I. There is a  $\bar{w} \in (0,1]$  such that an informative EPBE exists if  $w < \bar{w}$ , but does not exist if  $w > \bar{w}$ .
- II. The principal's expected utility in the best EPBE is strictly decreasing in  $w$  on the interval  $(0,\bar{w}]$ , and is lower and constant in  $w$  on the interval  $(\bar{w},1)$ .
- III. In any informative EPBE, the principal's expected utility is at least as high as it would be if one action was always chosen.

Part I says that an informative EPBE exists for small  $w$ , but may not exist for large  $w$ .<sup>18</sup> More importantly, part II says that an increase in  $w$  decreases welfare. Equation (2) implies that as emotions become more important, the agent gives the principal good rather than accurate news—recommends  $t = 1$ —in more states, decreasing ex ante welfare. Feelings therefore have a self-defeating aspect: a principal with stronger emotions gets poorer recommendations, even though knowing that he is basing his decision on better advice would not only improve physical outcomes, it would make him feel better on average as well. Nevertheless, as part III shows, the principal is still better off in an informative EPBE than if he chose the same action independently of  $s$ . If the opposite was the case, he would not be willing to follow the agent's advice to choose  $t = 1$ .

The source of the agent's failure to maximize the principal's ex ante expected utility despite sharing his interests is an *informational externality*. In EPBE, the choice to recommend  $t = 1$  in more states—which is optimal *given* how the principal interprets messages—feeds back into the inferences driving emotions: each of  $t = 0$  and  $t = 1$  induce lower anticipatory utility as a result.<sup>19</sup> Since the agent does not take the feedback into account, he overuses the  $t = 1$  recommendation from an ex ante point of view.<sup>20</sup>

#### IV.B. Equilibria When the Agent Chooses the Action

In many relevant economic situations, the agent does not make recommendations, but rather takes actions herself that both affect the principal's physical utility and are observable to

18. While Part II shows that the distortion in the choice of  $t$  increases with  $w$ , it is *not* in general true that for a sufficiently large  $w$ , an informative EPBE does not exist. If there is an  $s_{\text{lim}}$  such that  $E[s - l(s,1)|s > s_{\text{lim}}] = E[s|s < s_{\text{lim}}]$  and  $E[l(s,0)|s > s_{\text{lim}}] > E[l(s,1)|s > s_{\text{lim}}]$ , then an informative EPBE exists for any  $w < 1$ , and—letting  $s^c(w)$  denote the best informative EPBE for  $w$ — $\lim_{w \rightarrow 1} s^c(w) = s_{\text{lim}}$ . Intuitively, as  $w$  increases, the distortion in the action becomes so severe that suggesting  $t = 1$  no longer makes the principal feel better than recommending  $t = 0$ . Thus, the agent's incentive to recommend  $t = 1$  diminishes.

19. When the agent expands the range of states for which she recommends  $t = 1$ , she takes the best states in which she recommends  $t = 0$  and turns them into the worst states in which she recommends  $t = 1$ . Hence, both of these recommendations will be associated with lower average states of the world.

20. One way to understand why an increase in  $w$  decreases welfare is through these informational externalities. Since they operate through the emotions generated by the interpretation of the agent's message, when only actual outcomes matter, they are nonexistent. The more important is anticipatory utility, the more important are the externalities, so the further away welfare from the optimal.

him. This is the case for many government actions, for most top-down managerial decisions, for almost all orders and decisions in the military, and for some choices parents make on behalf of their kids. The current section modifies the model to account for this possibility. The key result that  $t = 0$  is chosen too rarely survives, with an intriguing additional possibility: in the unique EPBE action choice may be so distorted that ignoring all information and *always* choosing  $t = 0$  would yield higher welfare.

The setup of the model is identical to the previous one's, except that the agent chooses  $t$  instead of the principal. In an EPBE of this new game, the agent chooses  $m$  and  $t$  to maximize the principal's expected utility (taking into account what inferences he will draw from them), and he updates beliefs using Bayes' rule, knowing the agent's EPBE strategy. An EPBE is simple if for any  $m, t$  and  $m', t$  that induce the same anticipatory utility,  $m = m'$ .

In a simple EPBE the agent always sends the same message along with a given  $t$ : if there were messages that induced different anticipatory utilities, she would always prefer to send the most positive message. Since the principal can therefore only infer information from  $t$ , a choice of action by the agent has the same consequences as does the corresponding recommendation in a setting where the principal always follows recommendations and infers no information from them beyond the recommended action. Since Proposition 2 and equation (2) were derived in exactly such a setting, these results extend to the new version of the model. But since it is now the agent who chooses  $t$ , there is no incentive-compatibility constraint that the principal follow a recommendation to choose  $t = 1$ . These observations lead to the following characterization.

PROPOSITION 4 (EPBE when the Agent Chooses  $t$ ). Suppose that (A1) holds.

- I. In any informative EPBE there is an  $s^c \in (0, s^*)$  such that  $t = 0$  is chosen exactly when  $s \leq s^c$ ; and there is such an EPBE if and only if  $s^c$  satisfies equation (2).
- II. Either an informative EPBE, or an EPBE in which  $t = 1$  is chosen with probability 1, exists.

As when the principal chooses  $t$ , equation (2) implies that expected utility in the best informative EPBE (in the range where

one exists) is strictly decreasing in  $w$ . There is, however, an important possibility in the current version of the model that was previously ruled out by the principal's incentive-compatibility constraint. It may be that  $E[l(s,1)|s > s^c] > E[l(s,0)|s > s^c]$  for any EPBE cutoff state  $s^c$ , so that  $t = 1$  is the suboptimal action on average even for the states in which it is chosen. In this case, any informative EPBE yields lower expected utility than if  $t = 0$  was always chosen. Furthermore, always choosing  $t = 0$  is often not an EPBE. It is not an EPBE if  $t = 1$  is very costly for high  $s$ , and under the reasonable condition on off-EPBE beliefs that  $t = 1$  induces weakly higher anticipatory utility than  $t = 0$ , it is not an EPBE for *any* parameters of the model.<sup>21</sup>

In these situations, welfare is actually *decreased* by the agent's private information. If she did not observe  $s$ , the unique EPBE would clearly be to choose  $t = 0$  with probability 1. Intuitively, once the agent has information, she will select  $t = 1$  when that information is favorable. Since  $t = 1$  then indicates good news to the principal, the agent chooses this action too often.

To illustrate, suppose that an army's capability, on the perception of which morale hinges, affects whether it is worth putting up a fight at a city that is difficult to defend. For both morale and strategic reasons, commanders will decide to fight if they are confident of a victory. Because a withdrawal then indicates military difficulties and can consequently cause morale to collapse, commanders are forced to defend the city in most other cases as well. Even if they do so, however, morale is severely hurt by soldiers' very realization that maintaining it is probably the sole reason for fighting. Hence, committing to ignore information and always surrender the city would be better.

#### IV.C. Further Comparative Statics

To further investigate the circumstances under which the parties fail to maximize their joint expected utility, I study how this tendency depends on the central features of the model.

First, in a sense, expected utility *increases* in the cost of choosing the wrong action:

21. Under such a condition, the agent chooses  $t = 1$  for all  $s > s^*$ , where this action yields strictly higher physical utility and weakly higher anticipatory utility than  $t = 0$ .

PROPOSITION 5 (The Effect of the Cost of Mistakes on Welfare).

Under (A1), the following statements hold both when the principal and when the agent chooses  $t$ :

- I. For any  $\epsilon > 0$ , there is a  $k > 0$  such that if  $L(s) \geq k(s^* - s)$  for all  $s \leq s^*$ , then the principal's expected utility in any informative EPBE is greater than  $E[s] - \epsilon$ .
- II. Suppose that  $s$  is uniformly distributed and there is a  $k > 0$  such that  $L(s) = k(s^* - s)$  for all  $s$ . Then, if an informative EPBE exists, it is unique. In the range where an informative EPBE exists, the principal's expected utility is strictly increasing in  $k$ .

Part I says that if the mistake of choosing  $t = 1$  when  $t = 0$  is called for is extremely costly, the agent is led to almost maximize expected utility. Because mistakes are very costly, making the correct physical decision determines the choice of  $t$  for all  $s$  outside a small knife-edge region. By equation (2), even in this region the loss in physical utility is bounded by  $w/(1 - w)$ , so the overall distortion is also small. Because an increase in the cost of mistakes improves action choices more generally, there is a tendency for it to increase expected utility. But there is an effect acting in the opposite direction: in the region where  $t = 1$  is still mistakenly chosen, an increase in the loss from choosing it decreases expected utility. Part II of the proposition shows that the net effect is positive when  $s$  is uniformly distributed and  $L(\cdot)$  is linear—when an increase in the cost of mistakes is a proportional one.

Second, a crucial property of the setup is that the physical-utility-maximizing action is correlated with the principal's future physical utility. This feature follows from two assumptions: that the achievable level of physical utility varies with  $s$ , and that the optimal action varies systematically with achievable physical utility. In contrast to the former assumption, it may be the case that the correct action always brings physical utility to the same baseline level (which can be normalized to zero). For instance, citizens may not know the reason for an interruption in electricity supply, but they may still know that in any event the service can likely be fully restored soon. Because the agent then has no incentives to manipulate the

principal's beliefs about  $s$ , taking or recommending the optimal action in all states is part of an EPBE.

**PROPOSITION 6** (Optimality when Attainable Physical Utility Is State-Independent). Suppose that  $\max_{t \in [0,1]} h(s,t) = 0$  for all  $s \in [0,1]$ . Then, both when the principal and when the agent chooses  $t$ , there is an EPBE in which an optimal  $t$  ( $t \in \operatorname{argmax}_{t' \in [0,1]} h(s,t')$ ) is chosen for each  $s$ .

In contrast to the assumption that the optimal action varies systematically with achievable physical utility, there are choices that are equally optimal for principals with high future physical utility and low future physical utility. For example, whether a manager believes a new software she just became aware of is worthwhile for an employee to learn is often unrelated to the employee's future. Just as in Proposition 6, recommendations are then accurate.

Third, even if different levels of attainable physical utility call for different actions, communication may be facilitated by the principal's *ignorance* about the nature of the decision-making problem. With a reinterpretation, the principal's unawareness of a relationship between future physical utility and the optimal action is as if there was no such relationship. Since he is not able to derive meaningful conclusions from the agent's recommendation, she has no incentive to mislead him, and no distortion arises.<sup>22</sup> This indicates that increasing the principal's knowledge about a decision-making situation, which might allow him to then make inferences from the agent's recommendation, can actually make communication regarding the situation more difficult. For example, if a patient sent to a specialist for evaluation is told that hospitalization is necessary only if the results are really bad, upon being hospitalized he will feel terrible. As a result, it becomes more difficult for a physician to send him to the hospital.<sup>23</sup>

22. Complete ignorance is an extreme case. If the principal suspects, say, that  $t = 0$  may be optimal only when  $s$  is low, but is uncertain about this, she makes two inferences when the agent chooses recommends  $t = 0$ . She deduces that  $s$  might be low, but also that  $t = 0$  might be optimal in more states than she had previously thought. Because of the second conclusion, the inference from the first is blunted, making the agent less tempted to recommend a high  $t$ . The less the principal knows about the world, the weaker is the first conclusion, and the stronger is the second, and hence the more accurate are recommendations.

23. This of course does not mean that all kinds of knowledge make the communication problem more serious. For example, if the principal's ability to learn about the state of the world himself improves—so that his priors about  $s$  become less dispersed—the distortion in recommendations typically decreases.

The results in the current and previous sections indicate that communication may be most suboptimal for choices in which taking one action is clearly associated with worse states of the world, and the physical-utility dimension of the problem does not ubiquitously dominate the emotional dimension. Many of the applications mentioned in Section III fit this description.

#### *IV.D. The Value of Commitment*

This subsection discusses how it may be possible to increase welfare by decreasing or eliminating the agent's direction in choosing her recommendation or action.

If the agent can publicly commit to choosing or recommending  $t = 0$  exactly when  $s \leq s^*$ , the principal's expected utility is maximized. Some professional norms and guidelines may partly serve such a commitment role. In many situations, however, it is difficult to precisely describe in advance the circumstances in which a specific action is called for, and the value of an expert is often in being able to make the right judgment in exactly these contingencies. Norms then cannot work very well.

But even when commitment to state-contingent behavior is not feasible, commitment to always taking or recommending the same action might be. By Part III of Proposition 3, such commitment cannot increase welfare when the principal chooses  $t$ , but by the discussion in subsection IV.B, it can do so when the agent chooses  $t$ . This point can be illustrated with the threat advisory system run by the Department of Homeland Security, which very often warns of an elevated risk of terrorist attacks. A truly informative warning of an attack is likely to create so much fear in citizens that (by the results of subsection IV.B) officials would issue such a warning way too rarely. Partly to avoid this outcome, they instead issue a warning in even the mildest of situations—and publicly make a point of following this policy. Before September 11, when concerns about terrorism were not so great, the policy of constant warnings may not have been seen as appropriate, leading to an equilibrium in which any warnings were issued too rarely.

### V. SUPPLY OF CERTIFIABLE INFORMATION

The analysis so far was based on the assumption that the agent cannot directly communicate  $s$ . While this is the appro-

appropriate assumption for the majority of situations in applications, there are also circumstances where the agent can reveal or “disclose”  $s$ ; that is, supply her information to the principal. Sometimes, this is possible because the principal can verify the information; for instance, a doctor can often show objective test results (such as a blood-pressure measurement) to a patient. More often, this is possible because—either for moral reasons as in many personal relationships, or for legal reasons as in the case of some managerial decisions—the agent would not convey false information, so that even her cheap-talk statements can be trusted. The current section shows that in these cases, the agent supplies good news to the principal, but only reveals bad news if this is necessary to induce him to take the right action.

In contrast to the model introduced in Section II, which assumes that the agent observes  $s$  with probability 1, this section allows for the realistic and interesting possibility that she does not learn it. When  $s$  is not certifiable, this assumption does not make a qualitative difference to the results.

#### *V.A. Information Disclosure without Action Choice*

To simplify the exposition, I first investigate whether the agent chooses to supply information when the action choice is trivial:  $t = 0$ , and  $l(s, t) \equiv 0$ . The results extend and modify those of Caplin and Leahy [2004]. In subsection V.B, I reintroduce a nontrivial action choice.

The game is now the following. With probability  $\alpha \in [0, 1]$ , the agent privately observes  $s$ . If she does, she can send the disclosure message  $d$ , or send any cheap-talk message  $m \in \mathcal{M}$ . If she does not learn  $s$ , she can only send cheap-talk messages. This implies that she cannot certifiably communicate *not* having learned  $s$ . In situations where available information is hard, this assumption captures the notion that it is typically still impossible to certify that one does not have information. In relationships in which parties would for moral or legal reasons not falsify their information, the same assumption means that not providing available information is still acceptable.

Modifying the definition of EPBE to incorporate the possibility of sending a disclosure message is straightforward: the principal’s belief upon receiving  $d$  must assign unit mass to  $s$ . The

other conditions of EPBE, and of simple EPBE, are the same.<sup>24</sup> As before, I restrict attention to simple EPBE.

I first simplify the analysis with the following lemma.

LEMMA 3. Suppose that  $h(s,t) = s$  for all  $s$  and there is no action choice. Then, in any simple EPBE, the agent sends the same message whenever she does not disclose  $s$ .

Lemma 3 means that the nature of the communication game effectively reduces the message space to two options: either reveal  $s$  or remain silent. This leads to the following proposition.

PROPOSITION 7 (Information Disclosure). Suppose that  $h(s,t) = s$  for all  $s$  and there is no action choice. Then, there is a unique  $\bar{s}$  such that the agent, in case she observes  $s$ , discloses it if and only if  $s \geq \bar{s}$ . If  $\alpha = 1$ , then  $\bar{s} = 0$ ; and if  $\alpha < 1$ , then  $\bar{s} > 0$ .

The result for  $\alpha = 1$  is a generalization of Caplin and Leahy's [2004] full-disclosure result from two to a continuum of states. Since revealing news very close to the best possible induces nearly the highest possible anticipatory utility in the principal, the agent wants to disclose such news. Given that the principal knows this, no news would have to indicate to him that the state of the world is below the very best. Thus, the agent wants to reveal the best of those remaining states as well. Repeating the same logic shows that she reveals all information.

If  $\alpha < 1$ , however, the agent does not disclose all  $s$  she observes. When she "pretends" not to have bad news, in the principal's mind there remains a possibility that she may be uninformed, raising his anticipatory utility.<sup>25</sup>

### *V.B. Information Disclosure with Action Choice*

I now extend the disclosure model to nontrivial action choice. Thus, the game is the following. The agent observes  $s$  with prob-

24. Technically, one must also modify Definition 1 in the Appendix to account for the possibility that the agent does not know  $s$ . In this contingency, she maximizes the principal's expected utility according to her prior, not according to the actual state of the world. In addition, in the current model the principal's action choice is trivial, so the part of the definition pertaining to that choice is redundant.

25. Despite the nonstandard nature of the utility function in the model, both the formal result in and the intuition behind Proposition 7 are analogous to disclosure results in corporate finance. See, for example, Grossman and Hart [1980], Milgrom [1981], and Jung and Kwon [1988]. Bolton and Dewatripont [2005] provide a textbook treatment.

ability  $\alpha$ , and can either send a message  $m \in \mathcal{M}$  to the principal, or, if she has observed  $s$ , disclose it. Then, the principal chooses  $t \in \{0,1\}$ .<sup>26</sup>

The following proposition identifies a new type of EPBE in this case.

**PROPOSITION 8 (Information Disclosure with Action Choice).** Suppose that (A1) holds. If  $\alpha = 1$ , the agent reveals all  $s$  and the physical-utility-maximizing action is always taken. For  $\alpha < 1$ , the following type of EPBE exists. There is an  $\bar{s} \in (0,1)$  and a closed set  $N \subset [0,\bar{s}]$  such that (i) the agent discloses all  $s \geq \bar{s}$  and all  $s \in N$ , and does not disclose any other  $s$ ; (ii) she sends the same message whenever she does not disclose; and (iii) the principal chooses  $t = 0$  for all  $s \in N$ .

As the first statement in Proposition 8 indicates, for  $\alpha = 1$  the agent reveals all states of the world, and the appropriate action is always chosen. This result extends the full-disclosure case from Proposition 7, with a similar “unraveling” intuition: if the principal expected the agent not to reveal some states, she would want to disclose the best of those states, improving his beliefs about  $s$  and not making his choice of  $t$  worse.

For  $\alpha < 1$ , there is an interesting new type of EPBE, in which the states the agent reveals are not necessarily all better than the ones she hides. For example, it could be the case that she reveals good news, hides intermediate news, and also discloses bad news, with the bad news being transmitted only in states where  $t = 0$  is optimal. In this type of EPBE, if the agent does not reveal  $s$ , the principal chooses  $t = 1$  with positive probability. If  $s$  is low, such an action would substantially decrease physical utility, so the agent reveals  $s$  to make sure the principal realizes the gravity of his problem, and does the right thing. If  $s$  is somewhat higher, choosing  $t = 1$  is not such a great mistake, so the agent prefers to maximize anticipatory utility by not disclosing  $s$ . For example, if a spouse has a slight dissatisfaction with how irresponsible her husband is, she may shield his feelings by not telling him about it. But if his irresponsibility is so serious that it threatens the marriage, she may be blunt with him to make sure the problem is addressed.

26. In this section, I only analyze the model in which the principal chooses  $t$ . In the key applications where the agent chooses  $t$  (government policy, especially regarding the economy and security, as well as military and management decisions), relevant information generally cannot be communicated directly.

## VI. CONCLUSION

Throughout this paper, I have assumed that the principal is strategically sophisticated; he understands when the agent recommends or chooses  $t = 1$  versus  $t = 0$ . In many situations, it seems possible that he would take the agent's behavior at face value, naïvely assuming that she always recommends or chooses the action that maximizes physical utility. As her sophisticated counterpart, under (A1) a naïve principal also feels better if he thinks  $t = 1$  is the appropriate action. Hence, the result that action choice is slanted toward  $t = 1$ , and that this tendency increases in the weight  $w$  attached to anticipation, extends to a naïve principal. Because the principal does not account for such slanted choices in interpreting the agent's behavior, however, her behavior is *ex ante* optimal. This implies that commitment to a strategy is never valuable, and leads to different welfare conclusions than the sophisticated-principal model. In fact, the greater is  $w$ , the more important are the principal's unrealistically positive emotions, so the *greater* is his average total utility.<sup>27</sup> In addition, a naïve principal has higher welfare than a sophisticated one.

A natural extension of the model would be to consider dynamic situations, in which there are multiple opportunities for the parties to communicate. In such an environment, the *timing*—as opposed to merely the *content*—of information transmission is a new strategic consideration facing the agent. In the disclosure model, for instance, she may choose not to reveal bad information early on in the exchange. But as the principal becomes more and more skeptical that no news is available, or the time for choosing the right action draws near, she may reveal more news.

Finally, in this paper the existence of a relationship between the parties is assumed exogenously: the principal cannot select the agent he interacts with. Often, patients choose their doctors, citizens elect their governments, and employees select where to work. Identifying conditions under which reputational concerns or competitive forces mitigate or exacerbate communication problems is an important future agenda.

27. The principal's expected physical utility is still decreasing in  $w$ .

APPENDIX 1: DEFINITION OF EMOTIONAL PERFECT BAYESIAN  
EQUILIBRIUM

This section defines emotional perfect Bayesian equilibrium formally for the game introduced in Section II. Let  $\sigma_a(s, m)$  be the probability that the agent sends message  $m$  when the state of the world is  $s$ ,  $\sigma_p(m, t)$  the principal's probability of choosing action  $t$  after message  $m$ , and  $\mu(m)$  his beliefs about  $s$  after message  $m$ .

DEFINITION 1.  $\sigma_a(\cdot, \cdot)$ ,  $\sigma_p(\cdot, \cdot)$ , and  $\mu(\cdot)$  constitute an *emotional perfect Bayesian equilibrium* (EPBE) if

1. Agent optimization—For all  $s \in [0, 1]$ , if  $\sigma_a(s, m) > 0$ , then
 
$$m \in \operatorname{argmax}_{m' \in \mathcal{M}} \sigma_p(m', 0) \cdot U(\mu(m'), s, 0) + \sigma_p(m', 1) \cdot U(\mu(m'), s, 1).$$
2. Principal optimization—For all  $m \in \mathcal{M}$ , if  $\sigma_p(m, t) > 0$ , then

$$t \in \operatorname{argmax}_{t' \in \{0, 1\}} \int U(\mu(m), s, t') d\mu(m)(s).$$

3. Updating—For any  $m \in \mathcal{M}$  that is sent by the agent with positive probability,  $\mu(m)$  is obtained from the prior and  $\sigma_a$  using Bayes' rule.

An EPBE is *simple* if for all  $m, m' \in \mathcal{M}$ ,

$$\begin{aligned} \sigma_p(m, t) = \sigma_p(m', t) \text{ and } \max_{t \in \{0, 1\}} \int h(s, t) d\mu(m)(s) \\ = \max_{t \in \{0, 1\}} \int h(s, t) d\mu(m')(s) \Rightarrow m = m'. \end{aligned}$$

APPENDIX 2: PROOFS

Throughout this section, let the probability density and cumulative distribution functions of the random variable  $s$  be denoted by  $f(s) > 0$  and  $F(s)$ , respectively.

*Proof of Lemma 1.* We first show by contradiction that in any simple EPBE, no two messages that are sent with positive probability lead the principal to use the same strategy. If there were two such messages, they would have to induce the same antici-

patory utility; otherwise, the agent would never choose the one that induces lower anticipatory utility. But in that case, the EPBE would not be simple, a contradiction.

Therefore, we can order the agent's messages by the probability that the principal will choose  $t = 1$  after the message. Since  $L(s)$  is strictly decreasing, for any given anticipatory utilities induced by two messages, the relative utility of choosing the one that induces a higher likelihood of action  $t = 1$  strictly increases in  $s$ . Hence, the agent's message is an increasing step-function of  $s$  in the above order, so that each message is sent for an interval of states. Now we distinguish two cases, depending on whether there is a message that leads the principal to mix between the two actions.

Suppose that there is a message  $m$  sent by the agent with positive probability that leads the principal to mix. Then, the interval of states for which the agent sends  $m$  must contain  $s^*$  in its interior; otherwise, the principal could not be indifferent between  $t = 0$  and  $t = 1$ . Then, there is no message that leads the principal to choose  $t = 1$  with a higher probability: if there was such a message, it would be associated with higher states of the world, and hence the agent would prefer to send it for all  $s > s^*$  for both anticipatory and physical-utility reasons. Further, given that  $m$  induces the highest probability of  $t = 1$  and this probability is increasing in  $s$ , the agent sends  $m$  for all  $s > s^*$ . Therefore, any other message leads the principal to choose  $t = 0$  with probability 1. And, as noted above, in a simple EPBE there can only be one such message.

Now suppose that the principal follows a pure strategy after all messages sent with positive probability. Then, again from the above fact that in a simple EPBE such messages lead to different actions, there can only be two such messages.

*Proof of Proposition 1.* In text.

*Proof of Proposition 2.* Suppose that two messages sent with positive probability are  $m_0$  and  $m_1$ , and the principal chooses  $t = i$  with probability 1 after  $m_i$ . First, since  $L(s)$  is strictly decreasing, for any emotions induced by  $m_0$  and  $m_1$  the agent follows a cutoff strategy, sending  $m_1$  for states of the world above a given cutoff. Call this cutoff state of the world  $s^c$ .

Now we prove by contradiction that it cannot be an EPBE to have  $s^c \geq s^*$ . If this was the case,  $m_1$  would make the principal feel better than  $m_0$ :

$$\begin{aligned} \max_{t \in \{0,1\}} \int h(s,t) d\mu(m_1)(s) &= \int_{s^c}^1 s dF(s) > \int_0^{s^c} s dF(s) \\ &\geq \max_{t \in \{0,1\}} \int h(s,t) d\mu(m_0)(s). \end{aligned}$$

Therefore, for  $s = s^*$ , the agent, for a combination of instrumental and anxiety reasons, strictly prefers to send  $m_1$ . Since  $L(s)$  is continuous, she also prefers to send  $m_1$  for states slightly below  $s^c$ . This contradicts  $s^c \geq s^*$ .

*Proof of Lemma 2.* In text.

*Proof of Proposition 3.* For notational simplicity, define  $\kappa = w/(1-w)$ . Note that  $\kappa$  is increasing in  $w$ . On the interval  $[0, s^*]$ , consider the curves  $L(s)$  and

$$A(s) = E(s' - l(s',1)|s' \geq s) - E(s'|s' \leq s).$$

I. We first prove that if an informative EPBE does not exist for weight  $w$  attached to emotions, it also does not exist for any  $w' > w$ . By Lemma 2, if an informative EPBE does not exist for  $w$ , either there is no  $s^c$  such that  $\kappa A(s^c) = L(s^c)$ , or for any such  $s^c$ ,  $E[l(s,0) - l(s,1)|s > s^c] < 0$ .

First, suppose that there is no  $s^c$  such that  $\kappa A(s^c) = L(s^c)$ . Since  $A(s^*) > L(s^*) = 0$  and the two curves are continuous, this implies that  $\kappa A(s) > L(s)$  for all  $s \in (0, s^*]$ . Therefore, for any  $\kappa' > \kappa$ ,  $\kappa' A(s) > L(s)$  for all  $s \in (0, s^*]$ , so an informative EPBE does not exist for  $\kappa'$  and the corresponding  $w' > w$ .

Second, suppose that for all  $s^c$  satisfying  $wA(s^c) = (1-w)L(s^c)$ ,  $E[l(s,0) - l(s,1)|s > s^c] < 0$ . Then, this is in particular true for the highest  $s^c$  such that  $wA(s^c) = (1-w)L(s^c)$ . Call this highest cutoff  $s^h$ . By the proof of Part II of the proposition below, for any  $s^c$  satisfying  $w'A(s^c) = (1-w')L(s^c)$ , we have  $s^c < s^h$ , and hence  $E[l(s,0) - l(s,1)|s > s^c] < 0$ . Thus, once again an informative EPBE does not exist for  $w'$ .

To complete the proof of Part I, we prove that an informative EPBE exists for a sufficiently small  $w$ . Take a  $0 < s < s^*$  satisfying  $E[l(s',0) - l(s',1)|s' > s] > 0$ . By the continuity of  $l$ , such an  $s$  clearly exists. Since  $L(s) > 0$ , there is a  $w > 0$  such that  $wA(s) < (1-w)L(s)$ . Again using that  $A(s^*) > L(s^*) = 0$  and the two curves are continuous, there is an  $s^c \in (s, s^*)$  satisfying the conditions of Lemma 2.

II. We first prove the part of the statement pertaining to  $w < \bar{w}$ . Let  $s^c(w)$  be the cutoff in the best EPBE. Clearly,  $s^c(w)$  is given by the highest intersection of the curves  $wA(s)$  and  $(1 - w)L(s)$ . Since  $A(s)$  and  $L(s)$  are continuous and  $A(s^*) > L(s^*) = 0$ , we know that  $wA(s) > (1 - w)L(s)$  for any  $s > s^c(w)$ .

Now take any positive  $w$  and  $w'$  such that  $w' > w$ . Since  $A(s)$  and  $L(s)$  are continuous and  $wA(s^c(w)) = (1 - w)L(s^c(w)) > 0$ , there is some  $\epsilon > 0$  such that  $w'A(s) > (1 - w')L(s)$  for any  $s > s^c(w) - \epsilon$ . This means that  $s^c(w') \leq s^c(w) - \epsilon$ , strictly decreasing expected utility.

For any  $w > \bar{w}$ , the principal chooses an action that is optimal without any information. Thus, her expected utility is constant in  $w$ . By Part III of the proposition, this expected utility is lower than expected utility for the informative EPBE for  $w < \bar{w}$ .

III. Since  $s < s^*$  whenever  $t = 0$  is chosen, the principal's expected physical utility and by implication expected total utility is higher than if he always chose  $t = 1$ . And from the condition of Lemma 2 that in an informative EPBE  $E[l(s,0) - l(s,1)|s > s^c] \geq 0$ ,  $t = 1$  is the better option on average in the states in which it is chosen. Hence, the principal's expected utility is weakly higher than if he always chose  $t = 0$ .

*Proof of Proposition 4.* I. In text.

II. If an informative EPBE does not exist for weight  $w$  attached to emotions, then there is no  $s^c$  such that  $wA(s^c) = (1 - w)L(s^c)$ . Since  $A(s^*) > L(s^*) = 0$  and the two curves are continuous, this implies that  $wA(s) \geq (1 - w)L(s)$  for all  $s \geq 0$ . In particular,  $wA(0) \geq (1 - w)L(0)$ , so that  $wE[s - l(s,1)] \geq (1 - w)L(0)$ . Thus, we can construct an EPBE in which  $t = 1$  is chosen in the following way. Take some  $m_1 \in \mathcal{M}$ . The agent sends  $m_1$  for all states of the world, and always chooses  $t = 1$ . If she chooses  $t = 1$ , the principal's beliefs are equal to his prior. If she chooses  $t = 0$ , he believes that  $s = 0$  with probability 1. It is easy to check that this profile is a simple EPBE. The principal's inferences are clearly Bayesian, and by  $wE[s - l(s,1)] \geq (1 - w)L(0)$ , the agent will indeed want to choose  $t = 1$  for all  $s \geq 0$ .

*Proof of Proposition 5.* I. Let  $L^{-}(\cdot)$  be the inverse of  $L(\cdot)$ , which exists since  $L(\cdot)$  is strictly decreasing. We prove that for any  $\epsilon > 0$ , there is a  $\delta > 0$  such that if  $L^{-1}(w/(1 - w)) > s^* - \delta$ , in any informative EPBE the principal's expected utility is within  $\epsilon$  of the optimum  $E[s]$ . This implies the statement.

Since  $s \in [0, 1]$  and  $l$  is nonnegative everywhere, for any  $s^c < s^*$ , we have

$$(3) \quad w[E(s - l(s, 1)|s \geq s^c) - E(s|s \leq s^c)] \leq w.$$

Therefore, if  $L^{-1}(w/(1 - w)) > s^* - \delta$ , by equation (2) any informative EPBE cutoff  $s^c$  satisfies  $s^* - \delta < s^c < s^*$ . Using that  $l(s, 1) < w/(1 - w)$  for any  $s \in [s^c, s^*]$ , the loss in expected utility relative to the optimum is therefore bounded from above by  $w/(1 - w) \cdot (F(s^*) - F(s^c))$ . For a sufficiently small  $\delta$ , this is clearly less than  $\epsilon$ .

II. Consider again the two curves from the proof of Proposition 3, but index them according to  $k$ :  $L_k(s) = k(s^* - s)$  and  $A_k(s) = 1/2 - E[l_k(s', 1)|s' > s]$ , where  $l_k(s', 1) = k(s^* - s)$  for  $s \leq s^*$  and  $l(s', 1) = 0$  for  $s > s^*$ . Since  $L_k$  is strictly decreasing and  $A_k$  is strictly increasing, if there is an informative EPBE, it is unique. Let the cutoff state of the world in this informative EPBE be  $s^c(k)$ .

Now take  $k$  and  $k'$  such that  $k' > k$ . Since for  $s \in (0, s^*)$ ,  $L_{k'}(s) > L_k(s)$  and  $A_{k'}(s) < A_k(s)$ , we must have  $s^c(k') > s^c(k)$ . Now there are two cases.

First, suppose that  $L_{k'}(s^c(k')) \leq L_k(s^c(k))$ . In an informative EPBE with cutoff  $s^c$ , expected utility is equal to

$$E[s] - \text{prob}[s^c < s < s^*]E[l(s, 1)|s^c < s < s^*] = \frac{1}{2} - (s^* - s^c) \frac{L(s^c)}{2}.$$

Since  $s^* - s^c(k') < s^* - s^c(k)$  and  $L_{k'}(s^c(k')) \leq L_k(s^c(k))$ , expected utility is strictly higher for  $k'$ .

Next, suppose that  $L_{k'}(s^c(k')) > L_k(s^c(k))$ . This means that  $A_{k'}(s^c(k')) > A_k(s^c(k))$ , and so  $E[l_{k'}(s, 1)|s > s^c(k')] < E[l_k(s, 1)|s > s^c(k)]$ . Now using that expected physical utility with cutoff  $s^c$  is  $1/2 - \text{prob}[s > s^c]E[l(s, 1)|s > s^c]$  and that  $s^c(k') > s^c(k)$  and hence  $\text{prob}(s > s^c(k')) < \text{prob}(s > s^c(k))$ , the result follows.

*Proof of Proposition 6.* Take any two  $m_0, m_1 \in \mathcal{M}$  with  $m_0 \neq m_1$ . Let  $\sigma_a(s, m_0) = 1$  for  $s < s^*$ , and  $\sigma_a(s, m_1) = 1$  for  $s \geq s^*$ . That is, the agent sends message  $m_0$  when  $t = 0$  is appropriate, and  $m_1$  when  $t = 1$  is appropriate. Let the principal's strategy be to "follow" this advice:  $\sigma_p(m_i, i) = 1$  for  $i = 1, 2$ , and  $t = 0$  for all  $m \notin \{m_0, m_1\}$ . It is easy to construct beliefs that make these

strategies part of an EPBE. Namely, for any  $x \in [0,1]$ , let  $\mu(m_0)(x) = (\min\{F(x), F(s^*)\})/F(s^*)$  and  $\mu(m_1)(x) = (\max\{0, F(x) - F(s^*)\})/(1 - F(s^*))$ , and for any other message, beliefs attach probability 1 to the worst outcome.

*Proof of Lemma 3.* Suppose that  $m_1$  and  $m_2$  are messages sent by the agent with positive probability when she does not disclose  $s$ . Then, they induce the same anticipatory utility, since otherwise the agent would not want to send the one that induces the lower one. By assumption, the principal also follows the same strategy after the two messages. Thus, if we are in a simple EPBE,  $m_1$  and  $m_2$  must be the same message.

*Proof of Proposition 7.* We look for the EPBE value of  $v$ , the anticipatory utility of the principal when the agent does not reveal  $s$ . Clearly, for any given  $v$ , the agent reveals  $s$  if and only if  $s \geq v$ . For any  $v > 0$ , we can define

$$V(v) = \frac{\alpha F(v) E[s|s < v] + (1 - \alpha) E[s]}{\alpha F(v) + 1 - \alpha}.$$

$V(v)$  is the principal's anticipatory utility when the agent does not disclose  $s$ , and he believes that she discloses states of the world  $s \geq v$ . Hence, we have an EPBE in which anticipatory utility  $v > 0$  is induced if and only if  $v = V(v)$ .

Suppose first that  $\alpha = 1$ . Then for any  $v > 0$ ,  $V(v) < v$ , so  $v > 0$  is not compatible with EPBE. Clearly, there is an EPBE in which the agent discloses all  $s$ , and if she does not disclose, the principal believes that  $s = 0$  with probability 1.

Now consider  $\alpha < 1$ . In this case,  $V(\cdot)$  can be defined as above for any  $v \geq 0$ , and as before  $v$  is compatible with an EPBE if and only if  $v = V(v)$ . Notice that  $V$  is differentiable,  $V(0) > 0$ , and  $V(1) < 1$ . Thus, there is at least one EPBE. Now

$$\begin{aligned} V'(v) &= \frac{(\alpha F(v) + 1 - \alpha)\alpha v f(v) - (\alpha F(v) E[s|s < v] + (1 - \alpha) E[s])\alpha f(v)}{(\alpha F(v) + 1 - \alpha)^2} \\ &= \frac{\alpha f(v)(v - V(v))}{\alpha F(v) + 1 - \alpha}. \end{aligned}$$

Thus, whenever  $V(v) = v$ ,  $V'(v) = 0$ . This proves that there is a unique  $v^*$  such that  $V(v^*) = v^* > 0$ .  $\bar{s} = v^*$  satisfies the requirements of the proposition.

*Proof of Proposition 8.* Suppose first that  $\alpha = 1$ . We first establish that the agent reveals almost all states. Assume by contradiction that there is a set  $Q$  of states of measure greater than 0 that she does not disclose. Whenever the agent discloses the state, the principal chooses the physical-utility-maximizing action, so that his utility is  $s$ . When she does not disclose, his utility is  $w \cdot E[s' - l(s', t) | s' \in Q] + (1 - w) \cdot (s - l(s, t)) \leq w \cdot E[s' | s' \in Q] + (1 - w) \cdot s$ . Clearly, the set  $\{s | s \in Q, s > E[s' | s' \in Q]\}$  has positive measure. Since by revealing  $s$  in this set of states, the agent can increase the principal's utility, we have a contradiction.

Since the agent reveals almost all states, it must be the case that for all states she weakly prefers to reveal  $s$ ; if there was a state which she strictly preferred not to reveal, there would be a positive measure of such states. Finally, by assumption, the agent reveals all states where she is indifferent between revealing and not, so she reveals all states.

For  $\alpha < 1$ , let  $m_1$  be the message sent by the agent whenever she does not disclose  $s$ . We look for an EPBE pair  $v, p$ , where  $v$  is the principal's anticipatory utility when the agent does not disclose  $s$  ( $v = \max_{t \in \{0,1\}} \int (s' - l(s', t)) d\mu(m_1)(s')$ ), and  $p$  is the probability that he chooses  $t = 1$  in the same situation ( $p = \sigma_p(m_1, 1)$ ). (Since the agent always sends the same message when she does not disclose  $s$ , these numbers are well-defined.) Let

$$Q(v, p) = \{s | s < wv + (1 - w)[s - pl(s, 1) - (1 - p)l(s, 0)]\}.$$

$Q(v, p)$  is the set of states the agent chooses not to reveal if that induces anticipatory utility  $v$  in the principal and leads him to choose  $t = 1$  with probability  $p$ . Now define  $V : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}$  by

$$\begin{aligned} V(v, p) &= \frac{\alpha \text{prob}(s \in Q(v, p)) E[s - pl(s, 1) - (1 - p)l(s, 0) | s \in Q(v, p)] \\ &\quad + (1 - \alpha) E[s - pl(s, 1) - (1 - p)l(s, 0)]}{\alpha \text{prob}(s \in Q(v, p)) + (1 - \alpha)}, \end{aligned}$$

which is the principal's expected anticipatory utility if the agent does not disclose  $s$ , and he expects her not to disclose states in  $Q(v, p)$ . Also, define the correspondence  $T : \mathbb{R} \times [0, 1] \rightrightarrows \{0, 1\}$  by

$$\begin{aligned}
0 \in T(v,p) & \text{ if } \alpha \text{ prob}(s \in Q(v,p))E[l(s,1) - l(s,0)|s \in Q(v,p)] \\
& \quad + (1 - \alpha)E[l(s,1) - l(s,0)] \geq 0, \text{ and} \\
1 \in T(v,p) & \text{ if } \alpha \text{ prob}(s \in Q(v,p))E[l(s,1) - l(s,0)|s \in Q(v,p)] \\
& \quad + (1 - \alpha)E[l(s,1) - l(s,0)] \leq 0.
\end{aligned}$$

That is,  $T(v,p)$  is the set of actions the principal is willing to choose when he expects the agent not to disclose the states in  $Q(v,p)$ , and she does not disclose. Clearly,  $v,p$  defines an EPBE if and only if  $V(v,p) = v$  and  $0 \in T(v,p)$  if  $p = 0$ ,  $1 \in T(v,p)$  if  $p = 1$ , and  $T(v,p) = \{0,1\}$  if  $0 < p < 1$ . We prove that such an EPBE exists.

Clearly,  $V(1,p) < 1$ ; and for  $v = \min_{s \in [0,1]} s - pl(s,1) - (1 - p)l(s,0)$ , we have  $V(v,p) > v$ . Furthermore,  $V$  is differentiable in its first argument, and by a similar calculation as in the proof of Proposition 7,  $\partial V(v,p)/\partial v \leq 0$  whenever  $V(v,p) = v$ . Hence, for each  $p$  there is a unique  $v(p)$  such that  $V(v(p),p) = v(p)$ .

Since  $V(v,p)$  is continuous in  $p$ ,  $v(p)$  is continuous. If  $0 \in T(v(0),0)$ , we are done. If  $1 \in T(v(1),1)$ , we are done. Otherwise, since  $T(v,p)$  is upper semi-continuous, there is a  $p$  such that  $T(v(p),p) = \{0,1\}$ .

Finally, we prove that this EPBE satisfies the properties in the proposition. By construction, the agent always sends the same message whenever she does not disclose. Furthermore, notice that if the agent discloses some  $s \geq s^*$ , she also discloses any  $s' > s$ : if she discloses  $s$ , then  $ws \geq wv - (1 - w)(1 - p)l(s,0)$ , so  $ws' > wv - (1 - w)(1 - p)l(s',0)$  for any  $s' > s$ . This implies that the complement of  $Q(v,p)$ , which is exactly the set of states the agent discloses in EPBE, can be written as  $N \cup [\bar{s}, 1]$ , where  $N \subset [0, s^*]$ . Since  $Q(v,p)$  is open,  $N$  is closed. And since  $s < s^*$  for any  $s \in N$ , the principal chooses  $t = 0$  for any such  $s$ .

DEPARTMENT OF ECONOMICS,  
UNIVERSITY OF CALIFORNIA, BERKELEY

#### REFERENCES

- Bewley, Truman F., *Why Wages Don't Fall During a Recession* (Cambridge, MA: Harvard University Press, 1999).  
 Bluedorn, Allen C., "The Theories of Turnover: Causes, Effects, and Meaning," *Research in the Sociology of Organizations*, I (1982), 75-128.  
 Bolton, Patrick, and Mathias Dewatripont, *Contract Theory* (Cambridge, MA: MIT Press, 2005).

- Broughton, Philip S., "Government Agencies and Civilian Morale," *Annals of the American Academy of Political and Social Science*, CCXX (1942), 168–177.
- Caplin, Andrew, and John Leahy, "Psychological Expected Utility and Anticipatory Feelings," *Quarterly Journal of Economics*, CXVI (2001), 55–80.
- Caplin, Andrew, and John Leahy, "The Supply of Information by a Concerned Expert," *Quarterly Journal of Economics*, CXIV (2004), 487–505.
- Christakis, Nicholas A., *Death Foretold: Prophecy and Prognosis in Medical Care* (Chicago, IL: The University of Chicago Press, 1999).
- Crawford, Vincent P., and Joel Sobel, "Strategic Information Transmission," *Econometrica*, L (1982), 1431–1451.
- Damasio, Antonio R., *Descartes' Error: Emotion, Reason, and the Human Brain* (New York, NY: Quill, 2000).
- Dessein, Wouter, "Authority and Communication in Organizations," *Review of Economic Studies*, LXIX (2002), 811–838.
- Fang, Hanming, and Giuseppe Moscarini, "Morale Hazard," *Journal of Monetary Economics*, LII (2005), 749–777.
- Geanakoplos, John, David Pearce, and Ennio Stacchetti, "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, I (1989), 60–79.
- Grossman, Sanford J., and Oliver D. Hart, "Disclosure Laws and Takeover Bids," *Journal of Finance*, XXXV (1980), 323–334.
- Gurfein, Murray I., and Morris Janowitz, "Trends In Wehrmacht Morale," *The Public Opinion Quarterly*, X (1946), 78–84.
- Johnsrud, Linda K., and Vicki J. Rosser, "College and University Midlevel Administrators: Improving Their Morale," *Review of Higher Education*, XXII (1999), 121–141.
- Jung, Woon-Oh, and Young K. Kwon, "Disclosure When the Market Is Unsure of Information Endowment of Managers," *Journal of Accounting Research*, XXVI (1988), 146–153.
- Kőszegi, Botond, "Emotional Agency: The Case of the Doctor-Patient Relationship," Working Paper, 2004.
- Kydland, Finn E., and Edward C. Prescott, "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, LXXXV (1977), 473–492.
- Milgrom, Paul R., "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, XII (1981), 380–391.
- Miller, Delbert C., "Economic Factors in the Morale of College-Trained Adults," *American Journal of Sociology*, XLVII (1941), 139–156.
- Prendergast, Canice, and Lars Stole, "Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning," *Journal of Political Economy*, CIV (1996), 1105–1134.
- Ptacek, J. T., and Tara L. Eberhardt, "Breaking Bad News: A Review of the Literature," *The Journal of the American Medical Association*, CCLXXVI (1996), 496–502.
- Rose, Caroline Baer, "Morale in a Trade-Union," *American Journal of Sociology*, LVI (1950), 167–174.
- Scharfstein, David S., and Jeremy C. Stein, "Herd Behavior an Investment," *American Economic Review*, LXXX (1990), 465–479.
- Todaro, John F., Biing-Jiun Shen, Raymond Niaura, Avron Spiro III, and Kenneth D. Ward, "Effect of Negative Emotions on Frequency of Coronary Heart Disease (The Normative Aging Study)," *American Journal of Cardiology*, XCII (2003), 901–906.
- Tulsky, James, "How Can Clinicians Better Support Patients at the End of Life," *DukeMed Magazine* (Spring/Summer 2004), 44–45.
- Zeitz, Gerald, "Structural and Individual Determinants of Organization Morale and Satisfaction," *Social Forces*, LXI (1983), 1088–1108.