Emotional Agency: The Case of the Doctor-Patient Relationship

Botond Kőszegi, UC Berkeley*

January 2004

Abstract

This paper identifies an array of complications in doctor-patient communication that arise when the patient suffers from anxiety. I assume that the patient derives utility from health outcomes as well as the anticipation of the exact same outcomes, and that the doctor wishes to maximize the patient's utility. The doctor privately (and probabilistically) observes a diagnosis, which affects the optimal treatment. She then sends a message to the patient, who chooses a treatment. While formulated in terms of medical care, this formal model of "emotional agency" also applies to many other situations. If the doctor cannot certifiably convey her diagnosis, communication is endogenously limited to a treatment recommendation, which the doctor distorts toward the treatment that is optimal when the patient is relatively healthy. Paradoxically, more emotional patients get less useful recommendations, even though basing their treatment choice on better recommendations would make them less anxious. If the doctor can certifiably convey her diagnosis, she does so for good news, but unless she needs to "shock" the patient into behaving correctly, she pretends not to know what is going on when the news is bad. If the patient visits two doctors in a row, the second doctor reveals more information than the first one. Results from an original survey of practicing physicians confirm that doctors care about patients' emotions, and alter their recommendations and other communication in response to them.

*Conversations with George Akerlof, Robert Anderson, Steve Goldman, Jonathan Gruber, Theodore Keeler, Wei Li, Ted Miguel, Markus Möbius, Ben Polak, Matthew Rabin, Antonio Rangel, Luca Rigotti, Tanya Rosenblat, Emmanuel Saez, and Chris Shannon contributed greatly to this paper.

1 Introduction

Patients' fears and hopes—and the myriad other emotions aroused by medical news—are of central concern to health professionals. Practicing physicians believe that such emotions affect medical decisionmaking and physical outcomes, and express difficulty in revealing sensitive information to their patients (e.g. Ptacek and Eberhardt 1996), a challenge for which they receive little training or guidance (Christakis 1999). Feelings thus seem to add an important dimension of complexity to doctor-patient interactions, an issue that has received little attention from economists. Yet standard elements of the economic theory toolbox (models of asymmetric information and mechanism design) may be uniquely helpful in identifying some of the problems inherent in doctor-patient communication, as well as in using this understanding to improve the provision of health care through better communication guidelines.

This paper is a contribution to the first of these steps: it studies the predicament doctors can find themselves in when trying to communicate with emotional patients. I start from the following three important aspects of the doctor-patient relationship, all of which have precursors in economics, but which have not been integrated into a comprehensive analysis of this topic. First, the doctor (she) is often in possession of superior information about the patient's (he) state of health, information she might want to and be able to disclose to him. In a different context, such disclosure decisions are familiar to economists from corporate finance (Milgrom 1981, Grossman and Hart 1980, Jung and Kwon 1988). Second, the doctor needs to guide the patient in choosing a treatment, a feature shared with models of expert advice (Crawford and Sobel 1982, Scharfstein and Stein 1990, Prendergast and Stole 1996, and others). Third, crucially, information about his health affects the patient's emotions. This ingredient is less mainstream in economics, but Caplin and Leahy (2003) have shown their theoretical importance in a disclosure setting. In particular, they prove that concerns for the patient's emotions can induce a doctor to reveal all of her information, because "no news" is interpreted by him as "bad news" (making him feel worse).

The key assumptions in my model are intended to help isolate the effects of emotions particularly starkly. Throughout the analysis, I assume that there is no conflict between the doctor and the patient: her goal is to maximize his utility. The patient's utility, in turn, is composed of two parts:

future health outcomes, and current fears, which depend on beliefs about the *exact same* health outcomes. For a majority of the paper, both components take the expected utility form, and a weight w is attached to the latter, anticipatory term. Since the patient's preferences over health outcomes (and indeed his decisions in all individual decisionmaking problems) are independent of w, this parameter can naturally be interpreted as capturing differences in the importance of anxiety in "otherwise identical" patients.[1]

The two parties face the following stylized situation, on which all results in the paper are built. The doctor privately (and probabilistically) observes a diagnosis s, which affects the optimal treatment. After a round of communication, the patient chooses one of two treatments.

Although this paper deals largely with the medical care application, it is important to note that this same model applies to many other situations as well, making it a more general theory of agency with emotions. Advisors need to give advice to their graduate students on the quality of their ideas, but do not want to depress or discourage them by doing so. Parents play a crucial role in setting realistic expectations in their children, but also aim to maintain their confidence and interest in educational and athletic activities. Expressing various dissatisfactions with a lover can lead to meaningful change in a relationship, but hurting a partner's feelings is painful for everyone, and often undermines any improvement. And governments can prepare their citizens by informing them of upcoming economic, social, or environmental changes, but they do not want to create panic or anxiety in the population.

I consider two major categories of models, depending on whether the doctor can certifiably convey the diagnosis to the patient, and look for the perfect Bayesian equilibria of the resulting game in each case. For the first set of results, I assume that the doctor cannot certifiably communicate s. Then, I prove that in equilibrium she cannot communicate information other than what treatment she recommends. In addition, however, if one of the treatments tends to be optimal when the diagnosis is more favorable, she chooses her recommendation not only to optimize physical health,

[1] For presentational purposes, this paper will thus assume throughout that the doctor cares just as much about the patient's anxiety as he does. As I note below, the results would remain qualitatively unchanged if the doctor cared less about the patient's anxiety, or even if she cared exclusively about his physical health, but those physical outcomes depended on his emotions. It may also be the case that the doctor dislikes giving bad news to the patient simply because she finds the patient's emotional reaction difficult to handle.

but also to reassure the patient. Specifically, she distorts her recommendation toward the more "optimistic" treatment. For example, to inculcate a glimmer of hope, a doctor might too often recommend actively fighting an advanced form of cancer, even when physical health does not warrant such an intervention. This decision by the doctor may seem to be all for the better, but it leads to a kind of mindgame element that is at the heart of the paper. Namely, since the patient knows that the physician cares about his emotions, he discounts the positive signal involved in her choice of recommendation. Even worse, he also knows that she distorts her recommendation from a physical health point of view, and ultimately his emotions are exactly about his expectations regarding future health. Therefore, her attempts to make him feel better make him feel *worse* (on average) in the end! In other words, the doctor and the patient, whose joint goal is to maximize the patient's expected utility, fail at the task: they do not use the doctor's information optimally. And the higher is the weight attached to anxiety, the stronger are the doctor's incentives to manipulate the patient's beliefs, and therefore the worse are the health outcomes *and* fears. Thus, fears have a self-defeating aspect. In addition to formalizing this central problem, Section 3 analyzes mitigating factors. I show that when the patient is ignorant about medicine, or when choosing the wrong treatment is very costly, his expected utility is actually close to the optimum.

For the second set of results, I assume that the doctor can certifiably reveal her diagnosis if she has one, but still cannot certifiably show that she does *not* have a diagnosis (which could occur with positive probability). To simplify the exposition, I start by abstracting away from treatment choice, making my setup similar to that of Caplin and Leahy (2003), the only other paper that studies the doctor-patient relationship when anxiety influences the situation. They assume that the doctor observes, with probability one, one of two possible diagnoses, and that she can certifiably communicate the diagnosis. They prove a full disclosure result in that setting: the doctor reveals all of her information in equilibrium.[2] I generalize this result, and show that (similarly to the disclosure literature in corporate finance) if the probability of obtaining a diagnosis is less than one, the doctor reveals relatively favorable news, but she does not reveal bad information even if

[2] In addition to their setup, this paper also allows for non-trivial treatment choice, considers the possibility that the doctor does not observe the diagnosis, analyzes behavior when she cannot send certifiable messages, and touches on the dynamics of communication.

she does observe it. For example, if the lab analysis of a patient's samples indicated the presence of a malignant tumor, a physician may claim that she has not yet received the results, sheltering the patient's feelings. Next, I assume that there is a second doctor the patient visits, who is identical to the first one in all respects: she does not learn the first doctor's diagnosis, she probabilistically observes a diagnosis of her own, she communicates with the patient and shares his utility function. It turns out that the second doctor discloses a wider range of diagnoses than the first, even if she is not more informed (that is, even if she does not observe the first doctor's message). The reason is that the patient interprets two doctors' refusal to inform him more negatively than just the first one's, because he finds it less likely that neither doctor knows the diagnosis. This gives the second doctor more incentive to reveal s. Thus, the patient learns bad diagnoses with more delay than good ones.

The partial information revelation results in this part of the paper invite the natural question: do patients who prefer to know the truth receive more information than those who would rather not know? With one doctor, all types receive the same amount of information. With two doctors, however, the conclusion is that information-averse patients receive *more* information from the first doctor than information-loving ones. Information-aversion means that the patient is more hurt by bad news than he is pleased with good news. His dislike of negative information motivates the first doctor to shield him from any risk of what the second doctor might tell him. But her only way of shielding him is to break the news to him herself. As a result, in equilibrium she (and possibly the second doctor as well) end up telling more.

Finally, I reintroduce treatment choice into the picture. A new type of equilibrium now becomes possible, in which the doctor discloses good news, and also some very bad news, but does not reveal some diagnoses in-between. Intuitively, failing to inform the patient of very bad news would leave him too optimistic, and he would choose the wrong treatment. Thus, the doctor needs to "shock" him with the truth in order to make him behave correctly.

There is very little precise evidence in the medical literature on how doctors transmit sensitive information to patients. To supplement what little there is, I have conducted a survey of practicing U.S. physicians on doctor-patient communication. While this evidence should by no means be taken as conclusive, overall the results support the assumptions and predictions of the model. To start, doctors express a nearly unanimous concern for patient emotions, and do so in multiple ways: they believe that emotions are an important determinant of physical outcomes, and repeatedly mention a patient's emotional or psychological state as a key factor in determining how communication should be conducted with them. In addition, physicians stress that patients are quite adept at interpreting their recommendations and other information. Most interestingly, in a wide variety of situations, doctors distort their recommendations, and dampen or delay in relaying negative information, as this paper's model predicts. A regularity of independent interest is that more experienced female doctors, and less experienced male doctors, are most likely to dampen or delay in passing on negative information to their patients.

2 Setup

2.1 The Utility Function

I begin with a discussion of the agent's utility function, which is intended to capture classical health concerns, as well as two properties of the emotional side of patient well-being: that patients have fears about their medical condition, and that these feelings respond to information.[3] There are two periods, 1 and 2, and all utility is derived from an (additive) combination of actual and anticipated health outcomes in period 2. Thus, I start from the end. In period 2, the patient derives utility from the state of his health, which is a function $h(s,t)$ of a random variable s realized in period 1, and some action t taken in the same period. In this paper, s will be interpreted as a "diagnosis," and takes a value in the interval $[0,1]$, with a continuous probability distribution function $f(s)$ and cumulative distribution function $F(s)$. The variable t can be interpreted in at least two ways: it could be a treatment received from the doctor or a health-relevant lifestyle choice the patient makes. t can take two values, 0 or 1.[4] The paper will consider several specifications for $h(s,t)$.

In the first period, the patient derives utility from his anticipation of health in the second period. Naturally, this anticipation can only depend on his beliefs about future outcomes (instead

[3] For references, see Kőszegi (2002).
[4] The basic results of the paper would still hold with more than two treatment options.

of the actual outcomes). Specifically, I assume that the patient's anticipatory utility depends on his expected health in period 2 conditional on his information in period 1. Formally, if μ is the cumulative distribution function that captures the patient's beliefs about s, and t is the chosen treatment, then anticipatory utility takes the form

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

μ will be determined endogenously (in a way specified below) by the patient's environment, his behavior, and that of the doctor. I assume that u is strictly increasing. The shape of u captures the patient's intrinsic preferences regarding information. If u is concave, he is called (analogously to risk aversion) "information-averse;" if u is convex, he is "information-loving;" and if u is linear, he is "information-neutral."[5]

The patient is an expected utility maximizer over the enriched payoff space that includes anticipation, and his von Neumann-Morgenstern utility function is intertemporally additive:

$$U(\mu,s,t) = \underbrace{u\left(\int h(s',t)d\mu(s')\right)}_{\text{anticipatory utility}} + \underbrace{h(s,t)}_{\substack{\text{health}\\ \text{outcomes}}}. \tag{1}$$

For conceptual clarity, my model therefore separates anticipatory utility from actual health outcomes. In reality, emotions also significantly affect physical outcomes, and the same formalism captures that case as well.

That anxiety depends on expected health instead of some other function of the patient's beliefs is not crucial for the qualitative results of the paper, but it is practical for two reasons. First, this functional form facilitates the comparison of patients who have different informational preferences, but who would make the same decision *given* their information:

$$\text{argmax}_{t\in\{0,1\}} \int U(\mu,s,t)d\mu(s) = \text{argmax}_{t\in\{0,1\}} \int h(s,t)d\mu(s). \tag{2}$$

[5] By the law of iterated expectations, the agent's conditional expected health does not change on average when he receives information. That is, the expectation of his expected health after he receives information is just his expected health before he receives information. Therefore, by Jensen's inequality, his average utility decreases due to information if u is concave, and increases due to information if u is convex.

Second, assuming specifically that u is linear—as will be done for a large part of the paper—makes the patient in effect a standard expected utility maximizer, allowing comparison to that model.[6]

Most of this paper will assume that the health outcome function $h(s,t)$ has the following two features. First, the patient's attainable level of utility varies with the diagnosis s. Second, the optimal treatment varies systematically with the attainable level of health. To capture the first of these properties formally, let $h(s,t) = s - l(s,t)$, with $l(s,t) \geq 0$ for all s,t and $\min_t l(s,t) = 0$. Then, for a diagnosis s, the attainable level of utility is s. But if the patient's treatment or lifestyle is not appropriate for his symptoms, he can be worse off as captured by the loss function $l(s,t)$.

To capture the second property above, I assume that there is a cutoff value $s^* \in (0,1)$ below which $t = 0$, and above which $t = 1$, is optimal ($l(s,0) = 0$ iff $s \leq s^*$); also, $L(s) = l(s,1) - l(s,0)$ is decreasing and continuous. This property of the loss function amounts to assuming that $t = 1$ is better when the patient is relatively healthy, and $t = 0$ is better when he is relatively sick. The following two very different medical situations are both consistent with such a loss function:

1. A patient's blood sugar is measured as part of his physical. If it is in the healthy range (s is high), then he can live his life as he used to ($t = 1$). But if his blood sugar is too high (s is low), a change in lifestyle is warranted ($t = 0$).

2. A prognosis is made for a patient who is known to have cancer. If there is a reasonable hope of survival (s is high), aggressive treatment ($t = 1$) is called for. However, if he is in the terminal stage (s is low), the best option is to do nothing ($t = 0$).[7]

Of course, many medical situations do not fit into this framework. At the end of Section 3.2, I discuss how alternative specifications modify the results.

The relationship of this model to the other possible applications mentioned in the introduction is straightforward. Graduate students, children, lovers, and citizens derive utility from their beliefs

[6] A utility function that incorporates emotions into decisionmaking has interesting implications for individual behavior regarding information acquisition. Kőszegi (2003) explores some of these implications. However, the current paper deals with the more complex situation in which the doctor also makes strategic decisions that affect the patient's information.

[7] Notice that doing nothing corresponds to $t = 1$ in the first example, and to $t = 0$ in the second. Thus, the same action can have different meanings in different situations.

about their future prospects and ability at the task in question, and advisors, parents, partners, and governments often have information that is useful for selecting the best action to take. Finally, the optimal action is often strongly correlated with the individual's prospects: a student should drop ideas that are bad, a child should diversify interests if she is not good enough to make a living with the current one, a lover needs to take steps to improve if he is not good enough in some way, and citizens need to take preparatory measures if economic or natural disaster is on the way.

2.2 The Communication Game

I now introduce the doctor as an active decisionmaker into the picture. Much of the health economics literature on doctor-patient relationships (and their interaction with insurance companies) focuses on agency problems due to the private interests of the different parties (see Mcguire 2000 for a review). To abstract away from these problems—which would largely be orthogonal to those below—I assume that the doctor's utility function is exactly the same as the patient's. Also, I now assume that u is linear, although later I will consider other utility functions as well. The von Neumann-Morgenstern utility function of both the doctor and the patient is thus

$$U(\mu, s, t) = w \cdot \int (s' - l(s', t)) d\mu(s') + s - l(s, t).$$

Since utility is linear in beliefs, the patient is an expected utility maximizer in that she maximizes expected utility from health in any individual decisionmaking problem. The parameter w captures the importance of anxiety, and will allow us to compare more or less "emotional" patients who have "otherwise identical" preferences. While this simplification helps in stating the results clearly, it does not affect the key insights.

The paper's basic results do not depend on the doctor fully internalizing the patient's anxiety, either; it is sufficient that she cares about it to a certain extent. The evidence in Section 5 indicates that doctors do. Furthermore, since emotional states affect physical outcomes, even physicians who care exclusively about the patient's medical condition are in the end forced to take feelings into consideration as well. Finally, it is important to note that most of my results apply in any situation in which the doctor does not like to give news which the patient interprets as bad.[8]

[8] The doctor might be reluctant to communicate unfavorable information because she cares about patient fear,

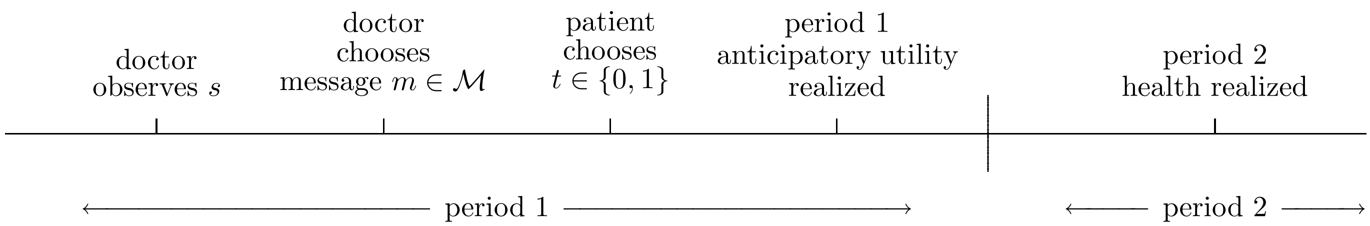

Figure 1: Timing

Given these claims of greater generality, an overarching feature of the paper's modeling strategy is worth highlighting. A major purpose of the above assumptions is not perfect realism or complete generality, but to isolate the role of emotions by *seemingly* eliminating any problems they may cause. Since anticipation ultimately depends on expected utility from physical outcomes, and the doctor has perfectly aligned interests with the patient, it might seem that feelings do not play a special role: taking best care of physical problems is equivalent to taking best care of feelings. Therefore, the paper's results follow from properties intrinsic to feelings, and not from any hidden arbitrary assumptions that emotions change preferences over physical outcomes or introduce conflicts between the two parties.

The doctor and the patient play the following game, with timing illustrated in Figure 1. First, s is drawn and is observed by the doctor. One can make two extreme assumptions regarding the doctor's ability to communicate s. The first half of the paper analyzes the case when she cannot certifiably reveal s itself. The second half of the paper assumes that she can reveal s. The former assumption is more plausible when a large part of the doctor's assessment is subjective or very difficult to quantify and communicate. This is certainly the case for lifestyle recommendations, but it is also the appropriate assumption for some treatable clinical conditions (e.g. depression; see Croghan 1991), and for most of the other applications of emotional agency. When the gist of the diagnosis is in the form of clear test results with unambiguous implications that the patient can understand, and forging the results is impossible or carries extreme sanctions, the latter assumption fits better.

but there might be other reasons as well: she might dislike the situation herself, or she might not want to deal with the patient's emotional reactions (Ptacek and Eberhardt 1996).

In this section, then, the game continues with the doctor choosing a cheap-talk recommendation m from a set of messages $\mathcal{M}$. For simplicity, I assume that $\mathcal{M}$ is finite. After observing m, the patient chooses the treatment or lifestyle t, and his anticipatory utility is calculated according to his beliefs after observing m and choosing t.[9]

In each of the cases to be discussed in the remainder of this paper, I look for perfect Bayesian equilibria of the game in Figure 1. In the present application, such an equilibrium needs to satisfy three requirements. First, the patient chooses the treatment optimally given his beliefs about s. Second, for any realized s, the doctor chooses her message m to maximize the patient's utility, correctly anticipating what inferences he will draw from m, and what treatment he will choose. (She cares about his inferences because they affect his emotions—and she cares about his emotions—and she cares about his choice of t because she cares about his health. And third, the patient uses Bayes' rule to infer the meaning of the doctor's message, knowing the doctor's strategy. Since the formal definition of perfect Bayesian equilibrium requires extra notation and is not necessary to understand the results, it is relegated to the Appendix.

Furthermore, a perfect Bayesian equilibrium is simple if two messages that lead the patient to choose the same treatment, and that induce the same expected health (and consequently the same anticipatory utility) in him, are in fact the same message. Notice that we can easily "transform" any perfect Bayesian equilibrium into a simple one by using a single message in place of any such equivalent messages.[10] Since the analysis in this paper centers around the patient's emotions and the distortions in treatment choice, this transformation changes nothing of interest. Thus, I restrict attention to simple perfect Bayesian equilibria for the rest of the paper.

[9] An alternative way of specifying the model is to assume that the doctor chooses t. I assume that the patient has this power because it seems slightly more realistic: although a doctor can recommend a treatment to the patient, she can rarely force anything on him. This is especially the case for lifestyle choices.

But ultimately it does not matter very much who chooses t. For any equilibrium with the patient choosing t, there is a corresponding equilibrium with the doctor choosing t which leads to the same outcome. When the doctor chooses t, there could be additional equilibria, but these equilibria are Pareto dominated by an equilibrium in which she always chooses $t = 0$.

[10] 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 patient, select some arbitrary $m \in M$. Then, change the strategy of the doctor so that whenever she would send a message in M, she sends m, adjust beliefs accordingly, and leave everything else unchanged. It is easy to check that this still satisfies the requirements of a perfect Bayesian equilibrium.

3 Optimistic Recommendations

3.1 Key Result

The current section explores behavior in the game introduced in Section 2. The following lemma drastically simplifies the analysis.

Lemma 1. *In any simple perfect Bayesian equilibrium, the doctor sends at most two messages with positive probability. If she sends exactly two messages with positive probability, one of the messages leads the patient to choose $t = 0$ with probability 1.*

As in other cheap-talk games, an uninformative or "babbling" equilibrium clearly always exists. In such an equilibrium, the doctor disregards s in choosing m, and the patient disregards m in choosing t. Similarly, an equilibrium in which the patient chooses $t = 0$ after one message and randomizes after the other one is effectively like an uninformative equilibrium, since it leads to the same expected health as choosing $t = 0$ regardless of the doctor's message. Thus, by Lemma 1, the only type of equilibrium in which the doctor's expertise actually helps the patient in choosing t is one where she sends two messages with positive probability, and he follows a pure strategy after each message. One can think of the two messages in informative equilibria as "I recommend $t = 0$" and "I recommend $t = 1$," with the patient following the doctor's advice. Thus, the nature of the situation endogenously limits the information the doctor is able to convey to the treatment she recommends. Intuitively, if she could communicate more information about the patient's futute health, she would always be tempted to give him the best possible news, leaving his choice of treatment unaffected and making him feel better. This may be one reason for the frustration patients sometimes feel in trying to get more precise information about their prospects out of doctors.

I now derive a key property of these "informative equilibria," and return to whether one exists later. Clearly, the ex ante optimal policy (and the one the doctor would commit to if she could) is to recommend $t = 0$ if and only if $s \leq s^*$. This is what maximizes the patient's expected health in the second period, and thus (by the law of iterated expectations) also his expected anticipatory feelings in the first period.[11] However, with the doctor free to choose her recommendation after

[11] The description of the ex ante optimal policy would be more complicated if u was non-linear. But the key result

observing s, it is not a perfect Bayesian equilibrium for her to follow this optimal policy:

Theorem 1. *In any informative equilibrium, there is an $s^c < s^*$ such that $t = 0$ is chosen for $s < s^c$, and $t = 1$ is chosen for $s > s^c$. Thus, $t = 0$ is chosen too rarely.*

To gain intuition for this result, suppose the patient believes that the doctor follows the ex ante optimal policy, and she observes $s \approx s^*$. In that case, from a physical health point of view, she is approximately indifferent as to what she should recommend, and her decision comes down to a consideration of the patient's anxiety. And in that respect, she strictly prefers to recommend $t = 1$, because it imparts the good news that $s > s^*$. In equilibrium, this motivation leads the doctor to expand the range of diagnoses for which she recommends $t = 1$.

Thus, in the presence of patient anxiety, the doctor cannot maximize the patient's expected utility, even though that is exactly what she is trying (and has the information) to do! Such an outcome is impossible with standard preferences: if two players in a game have exactly the same utility function, the strategy profile that maximizes their joint payoff (subject to informational constraints) is always an equilibrium. In particular, Crawford and Sobel (1982) present a standard model with a communication structure not unlike mine. In their setup, a well-informed sender sends a cheap-talk message to a receiver, who then takes an action that affects the utility of both players. But in sharp contrast to the current model with anxiety, if the parties' interests are perfectly aligned, they can use the sender's information optimally.[12]

In a certain sense, the doctor cannot maximize the patient's expected utility exactly *because* she wants to do so. In this model, if she cared only about the patient's physical health in period 2, she would not be motivated to reassure the patient, and would maximize his ex ante expected utility. Interestingly, then, she can maximize the patient's expected utility, but only if her objective is something different. Even so, since in reality physical health is affected by feelings (see Section

below—that the doctor recommends $t = 1$ too often relative to the ex ante optimal policy (Theorem 1)—would still be true in that case. With a non-linear u, there is no obvious natural way to state Corollary 1, because changing the weight on anxiety in general also changes the ex ante optimal policy.

[12] The consequences of Theorem 1 for other applications mentioned in the introducation are obvious. Advisors tell their students to drop ideas too rarely, parents are too reluctant to deflate their kids' expectations, lovers are too soft on pointing out each other's mistakes, and governments announce upcoming economic trouble less often than optimal.

5 and Kőszegi (2002)), the two kinds of outcomes are inherently inseparable, so requiring doctors to focus exclusively on physical outcomes does not eliminate the problem.

For $s^c < s^*$ to be the cutoff symptom in an informative equilibrium, the doctor needs to be indifferent between $t = 0$ and $t = 1$ when her diagnosis is s^c. This is equivalent to the condition

$$w\left(E[s - l(s,1)|s \geq s^c] - E[s|s \leq s^c]\right) = L(s^c). \tag{3}$$

Equation (3) can be used to address the question of the existence of informative equilibria. Such an equilibrium exists if and only if the highest s^c compatible with Equation (3) satisfies the condition that the patient actually wants to follow the doctor's recommendation. When the recommendation is $t = 0$, this is clearly the case, because in that range $s < s^*$, and so $l(s,1) > l(s,0)$. However, when the recommendation is $t = 1$, by (2) the patient will be willing to follow it if and only if $E[l(s,0) - l(s,1)|s > s^c] \geq 0$. Furthermore, when this condition holds, the best informative equilibrium is clearly better than any equilibrium in which one treatment is always chosen. It is easy to see that for a sufficiently small w, an informative equilibrium exists, and if it does not exist for some w, it also does not exist for any $w' > w$.

The right-hand side of Equation (3) is monotonically decreasing by assumption. However, there are no natural assumptions that would put restrictions on the left-hand side. Therefore, this game can exhibit multiple informative equilibria. Since the doctor and the patient share a common utility function, it may be natural for them to coordinate on the best perfect Bayesian equilibrium. This is the informative equilibrium with the highest s^c. The following corollary compares the welfare of "emotional" patients (for whom w is large) and their non-emotional counterparts in these best informative equilibria.

Corollary 1. *The patient's expected health in the best informative equilibrium is strictly decreasing in w.*

The intuition for Corollary 1 is closely related to that of Theorem 1: the doctor's incentive to manage the patient's emotions leads to a distortion in recommendations. As w increases, it becomes more important to reassure the patient, increasing the distortion. Fears thus have a self-fulfilling aspect: a more emotional patient gets poorer treatment on average, even though better treatment would not only improve physical outcomes, it would make him feel better as well.[13]

Although Theorem 1 and Corollary 1 demonstrate a failure of utility maximization, it bears emphasizing that the doctor's behavior *is* optimal ex post: given the patient's beliefs, she is choosing the best possible recommendation for each s. The key to understanding the result is that due to *informational externalities* between states, a strategy that is optimal for each s does not integrate into a strategy that is optimal ex ante. More precisely, by deciding to recommend $t = 1$ in more states, the doctor exerts an externality on the *interpretation* of the two recommendations for all other diagnoses s: it makes recommending both $t = 0$ and $t = 1$ a worse signal. The doctor, knowing s, does not care about the externality, but the patient (or the doctor ex ante), who is still ignorant, does. This leads the doctor to overuse the $t = 1$ recommendation.

One way to think about the difference between "cold" patients (who only care about actual outcomes) and "emotional" ones (who derive hedonic utility from their beliefs) is exactly to understand these cross-state externalities. The externalities operate through the interpretation of the doctor's message, so when only actual outcomes matter, they are non-existent. As more utility is derived from beliefs directly, the externalities become more and more important, and so the outcome is further and further away from the optimal.[14]

3.2 Mitigating Factors

To gain further insight into the kinds of medical situations in which this section's key effect—that the doctor makes optimistic recommendations too often—is likely to arise, I discuss several

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

[14] This way of understanding the problem also highlights that standard applications of expected utility theory force a specific interpretation of the preferences modelers assume. Though superficially compatible with both anticipation and caring about physical outcomes, the way expected utility theory is *used* is *only* compatible with the *cold* notion of decisionmaking. For example, it is standard procedure to decompose the set of possible states into different events the agent could find himself in, and find the best available action conditional on each of these events. The strategy thus obtained is thought to maximize the decisionmaker's expected utility. Unfortunately, in a multiplayer setup, this is only true if there are no cross-state externalities.

cases in which it is weakened or eliminated. First, a crucial property of my setup is that optimal treatments serve as signals of the patient's future health. This fact follows from two assumptions: that the patient's attainable level of health varies with s, *and* that the optimal treatment varies systematically with the attainable level of health. Either of these can be violated. In contrast to the former assumption, imagine that the patient visits the doctor with cold symptoms, which could be caused by a number of illnesses that are all perfectly treatable. This possibility can be captured by assuming that $h(s,t) = -l(s,t)$. Then, recommending the optimal treatment in all states of the world is part of a perfect Bayesian equilibrium.

Lemma 2. *If $h(s,t) = -l(s,t)$, there is a perfect Bayesian equilibrium in which the patient chooses the optimal t ($t \in argmin_{t' \in \{0,1\}} l(s,t')$) for each s.*

Intuitively, since the recommendation does not affect the patient's anticipatory utility, the doctor has no temptation to mislead him about the optimal treatment, and no distortion arises. In contrast to the assumption that optimal treatment varies systematically with the prognosis, there are recommendations (such as a no alcohol policy during pregnancy) that doctors are equally likely to give to healthy and unhealthy patients. Just as in Lemma 2, there is then no reason to distort the recommendation.

Second, even if different levels of potential health call for different treatments, the doctor and the patient may be able to take advantage of the patient's *ignorance*. With a reinterpretation, this is as if there was no relationship between the patient's level of health and the optimal treatment: rather than there being no relationship, the patient is just not aware of it. Once again, the doctor's recommendation then carries no anticipation-relevant information to him. Quite simply, an ignorant patient has to take a recommendation at face value, and will not be able to read anything else into it. Therefore, the doctor has no incentive to mislead him.[15] This indicates that providing patients with education about the attributes of their condition—which allows them to then make inferences

[15] Complete ignorance is an extreme case. In general, a less than fully educated patient makes two inferences when the doctor chooses to recommend, say, a low t. She partially deduces that s might be low, but she also learns about the appropriate treatments for this illness. Because of the second conclusion, the inference from the first is blunted, and thus the doctor is less tempted to recommend a high t. And the less the agent knows about medicine, the weaker is the first conclusion and the stronger is the second.

from doctors' recommendations—can actually make subsequent communication more difficult.[16]

Third, the importance of choosing the right treatment from a physical health point of view affects the distortion in the doctor's recommendation. In fact, the distortion disappears in the limit as L becomes very steep in the relevant range.

Lemma 3. *As $L^{-1}(w)$ converges to s^*, the patient's expected health in any informative perfect Bayesian equilibrium converges to the optimum $E[s]$.*

Intriguingly, if a mistake—in particular, the mistake of recommending $t = 1$ when $t = 0$ is called for—is extremely costly, the doctor is led to almost maximize the patient's expected utility. Intuitively, since the optimal treatment is very sensitive to the diagnosis, making the correct physical decision dominates the choice of t for all s outside a small knife-edge region. Noting that even in this region the maximum distortion for each s is on the order of w, the overall distortion is also small.[17]

The above arguments indicate that this section's effect may be strongest with chronic conditions that require a painful change in lifestyle, such as the example of high blood sugar used to motivate the model. For many of these conditions, having to change one's life is indeed a bad sign about health. Also, delaying the adjustment for a while may not result in a serious risk to one's health, so the loss function is not very steep.[18]

[16] This of course does *not* mean that all kinds of health education make the communication problem more serious. For example, if education improves the patient's ability to diagnose himself—so that his priors about s are less dispersed—the distortion in recommendations is in fact decreased. Thus, the overall effect of health education will depend strongly on the type of knowledge patients acquire. A fuller analysis of this issue is left for future work.

[17] The doctor's financial interests might also be aligned so that they counteract the communication problem due to emotions. If the recommendation $t = 0$ leads the patient to seek more of the doctor's services, while this is not the case with $t = 1$, the doctor will recommend $t = 0$ more often than predicted by this model. Of course, in one of the motivating examples above, when $t = 0$ was interpreted as a change in lifestyle, and $t = 1$ meant no change, this effect is non-existent. However, in many cases the worse news will indeed induce the patient to seek more medical help.

[18] At first sight, it may seem that if the loss function is not very steep, the patient's expected utility loss due to the distortion in recommendations cannot be too large. This is in general false: although in that case the mistake of recommending $t = 1$ increases slowly as s decreases, the doctor is at the same time more likely to be in the region where her recommendation is wrong. In fact, as the limit result of Lemma 3 examplifies, the opposite is closer to the truth.

4 Providing Information

The analysis of Section 3 was based on the assumption that the doctor cannot reveal the diagnosis s to the patient. The current section studies her behavior under the opposite assumption.

For a large part of this section, I abstract away from the need to choose a treatment, focusing on the information provision role of the doctor. This allows me to demonstrate most of the new effects in the cleanest possible manner. At the end of the section, however, I show how reintroducing a non-trivial treatment choice changes the results.

In addition to ignoring treatment choice for much of the analysis, I make another change relative to the model of Section 2. Namely, instead of assuming that the doctor observes the diagnosis s for sure, I allow for the possibility that she does not learn it. This makes the informational environment more realistic (and more interesting). Of course, this assumption is also very plausible when s is not certifiable. But in that setting, assuming that the doctor does not always observe the diagnosis would not make a qualitative difference to the results.[19]

4.1 Information Disclosure with One Period

The game being played by the doctor and the patient is now the following. With probability α, the doctor privately observes s. If she does, she can send the disclosure message d, or send any message $m \in \mathcal{M}$. If she does not learn s, she can only send a message $m \in \mathcal{M}$. In particular, this implies that she cannot certifiably communicate not having learned s.

Modifying the definition of perfect Bayesian equilibrium (Definition 1) to incorporate these changes is straightforward. To capture that sending the message d is disclosure, the belief it induces in the patient has to assign unit mass to s. Also, if the doctor does not know s, her maximization problem is slightly different than in Definition 1, since she averages the patient's utility over her own priors. The other conditions of perfect Bayesian equilibrium, and of simple perfect Bayesian equilibrium, are the same. As before, I focus attention on simple perfect Bayesian equilibria. In addition, as mentioned above, I ignore the choice of treatment for now (assuming that the same

[19] Generically, in a simple perfect Bayesian equilibrium, the doctor still sends at most two possible messages, and tends to distort towards the more "optimistic" recommendation.

treatment is always chosen and $l(s,t) \equiv 0$), and reintroduce this consideration only in Section 4.3.

I first simplify the analysis with the following lemma.

Lemma 4. *In any simple perfect Bayesian equilibrium, the doctor sends the same message whenever she does not disclose s.*

Lemma 4 allows us to think of the doctor's choice set as consisting of only two options: either reveal s or remain silent. Thus, as in the previous section, the nature of the problem effectively reduces the message space. This leads to the following theorem:

Theorem 2. *Suppose the doctor can disclose s. Then, there is a unique $\bar{s}$ such that the doctor, in case she observes s, discloses it if $s > \bar{s}$, and does not disclose it if $s < \bar{s}$. Furthermore, $\bar{s}$ satisfies*

$$\bar{s} = \frac{\alpha F(\bar{s}) E[s|s < \bar{s}] + (1-\alpha) E[s]}{\alpha F(\bar{s}) + (1-\alpha)}. \tag{4}$$

In particular, if $\alpha = 1$, the doctor discloses all diagnoses $s > 0$.

The result for $\alpha = 1$ is a generalization of Caplin and Leahy's (2003) "no news is bad news" effect. It is clear that the doctor would want to reveal news that are very close to the best possible. Given that the patient knows this, no news indicates that the diagnosis is below the very best. Thus, the doctor wants to reveal slightly worse news as well. The same logic applies repeatedly until she reveals all information.

However, if $\alpha < 1$, the doctor does not always disclose the diagnosis even when she does learn it. The intuition is that the doctor can use the possibility that she has not been able to come up with a diagnosis as an "excuse" behind which to hide when the news is bad. When she "denies" the existence of a diagnosis, the patient can harbor the illusion that his health could be the average $E[s]$. Although he knows that she might be withholding information—and adjusts his beliefs accordingly—for sufficiently low levels of s she will still choose to do so.[20]

[20] In the model, the doctor decides to hide a single piece of information that fully describes the patient's health. This should not be interpreted too literally. More often than not, real situations are not that simple; doctors might not be sure about the diagnosis themselves, and they may know more than one thing that could affect the patient's emotions. Nevertheless, the force of Theorem 2 (and of Theorem 3 and Corollary 2 below) applies. Doctors may hint at good news they are not sure about, but not at bad news, and they may hide some of their bad information while revealing other things.

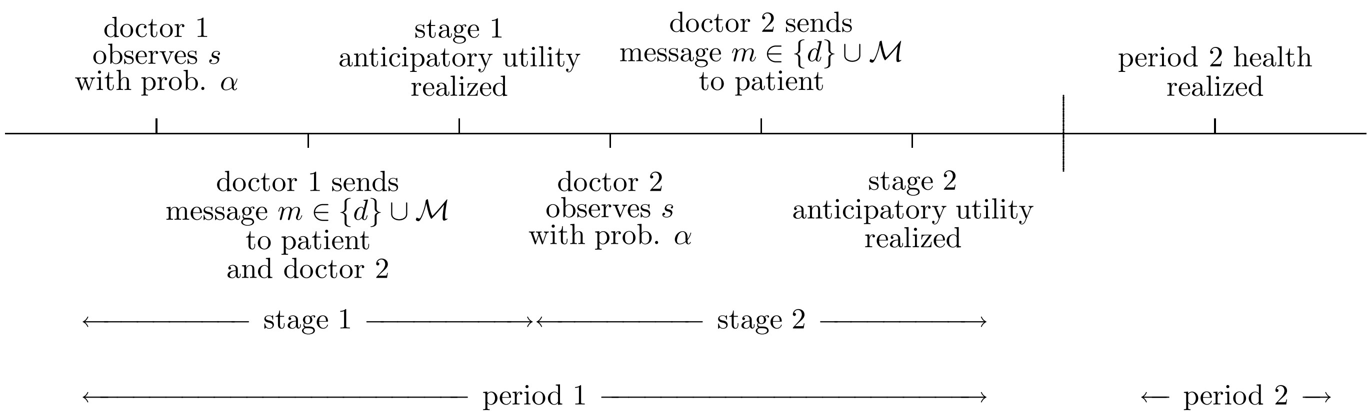

Figure 2: Timing of the Information Disclosure Model with Two Doctors

Despite the non-standard nature of the utility function in my model, both the formal result and the intuition in Theorem 2 are analogous to various disclosure results in corporate finance.[21] Although (for the case $\alpha < 1$) the result has not been applied to the doctor-patient framework, this still makes Theorem 2 more of an application than an original result.

Given that information disclosure is partial in Theorem 2, we can ask the natural question: how much information do patients with different informational preferences receive? Although the equilibrium Condition (4) applies to information-neutral agents, it is easily adapted for patients with different u's:

$$u(\bar{s}) = u\left(\frac{\alpha F(\bar{s})E[s|s < \bar{s}] + (1-\alpha)E[s]}{\alpha F(\bar{s}) + (1-\alpha)}\right).$$

Since u is strictly increasing, this is equivalent to Equation (4). Therefore, the amount of information the doctor reveals in equilibrium is independent of the patient's informational preferences. The reason is that for any given s, the doctor's disclosure decision depends not on the *shape* of u (on which informational preferences depend), but on a comparison of the *levels* of u with and without disclosure.

4.2 Information Disclosure with Two Doctors

I now consider a natural extension of the disclosure model, in which the patient visits two doctors in a row.[22] Suppose that there are two stages in period 1, in each of which the patient realizes anticipatory utility according to his most recent information. These components of utility enter additively and with equal weight in the patient's utility function, which both doctors share. The timing of the model is illustrated in Figure 2. Both doctors arrive at a diagnosis independently with probability α, and the second doctor does not learn the first doctor's diagnosis. At the beginning of stage 1, the first doctor communicates with the patient, and at the beginning of stage 2, the second doctor does. I also assume that the second doctor observes what the first doctor told the patient, but as I note below, the results would be identical if the opposite was the case (and so the results do not depend on her having this extra information).[23] A simple perfect Bayesian equilibrium is defined as a situation where no pair of messages by the two doctors induce the same anticipatory utility in the patient in both stages.

Lemma 5 in the Appendix shows that, with the above assumptions, the analogue of Lemma 4 holds: in any simple perfect Bayesian equilibrium, each doctor always sends the same message conditional on not revealing s. Thus, again we can think of the doctors' decisions as a choice between disclosing s and not doing so. Lemma 6 in the Appendix shows that doctors 1 and 2 follow cutoff strategies, disclosing $s > \bar{s}_1$ and (if the first doctor did not disclose) $s > \bar{s}_2$, respectively. Theorem 3 relates $\bar{s}_1$ and $\bar{s}_2$.

Theorem 3. *The perfect Bayesian equilibrium values of $\bar{s}_1$ and $\bar{s}_2$ are unique. Furthermore, we have $\bar{s}_1 > \bar{s}_2$, and $\bar{s}_1$ and $\bar{s}_2$ satisfy*

$$\bar{s}_2 = \frac{(1-\alpha)^2 E[s] + \alpha F(\bar{s}_2) E[s|s<\bar{s}_2] + \alpha(1-\alpha)F(\bar{s}_1)E[s|s<\bar{s}_1]}{(1-\alpha)^2 + \alpha F(\bar{s}_2) + \alpha(1-\alpha)F(\bar{s}_1)} \tag{5}$$

$$2\bar{s}_1 = \frac{(1-\alpha)E[s] + \alpha F(\bar{s}_1) E[s|s<\bar{s}_1]}{(1-\alpha) + \alpha F(\bar{s}_1)} + \alpha \bar{s}_1 + (1-\alpha)\bar{s}_2. \tag{6}$$

[21] For example, Grossman and Hart (1980), Milgrom (1981), and Jung and Kwon (1988). See Bolton and Dewatripont (2000) for a textbook treatment.

[22] Once again, this extension is more interesting when s is certifiable than when it is not. In Section 3, outcomes would be identical if two doctors in a row communicated with the patient.

[23] Note that I assume in this model that the patient always visits the second doctor. This is plausible, for example, when he needs to visit her to get treatment anyway. But the results would be the same if he visited her only in case the first doctor did not inform him.

Thus, the second doctor reveals more information—in particular, more unfavorable information—to the patient than the first one. In other words, bad news are revealed to patients with more delay than good news. The reason is that if the patient fails to receive a diagnosis from *both* doctors instead of just one, he finds it less likely that they do not have a diagnosis, and more likely that they are withholding it, making him feel worse. This induces the second doctor to provide more information. As this intuition suggests, Theorem 3 does not rely on the second doctor being more informed than the first. In fact, the statement and proof of the theorem would remain valid if the second doctor did not observe the first one's message.

I now return to the natural question: in equilibrium, do information-averse patients receive less information than information-neutral and information-loving ones? As far as information from the first doctor is concerned, exactly the opposite is the case. Recall Condition (6) for the first doctor's cutoff diagnosis $\bar{s}_1$ (above which she reveals s), and rewrite it for cutoffs $\bar{s}'_1, \bar{s}'_2$ when the patient has non-neutral informational preferences:

$$2u(\bar{s}'_1) = u\left(\frac{(1-\alpha)E[s] + \alpha F(\bar{s}'_1)E[s|s < \bar{s}'_1]}{(1-\alpha) + \alpha F(\bar{s}'_1)}\right) + \alpha u(\bar{s}'_1) + (1-\alpha)u(\bar{s}'_2). \tag{7}$$

Consider the case of an information-averse patient, for whom u is concave. For $\bar{s}'_1 = \bar{s}_1$ and $\bar{s}'_2 = \bar{s}_2$, Equation (7) is not satisfied: by Jensen's inequality, the left-hand side is strictly greater than the right-hand side. Therefore, the first doctor wants to reveal the diagnosis $\bar{s}_1$, in equilibrium leading to $\bar{s}'_1 < \bar{s}_1$. A mirror argument works for an information-lover.

Corollary 2. *If u is strictly concave, then for any $\bar{s}'_1$ satisfying Condition 7, $\bar{s}'_1 < \bar{s}_1$. If u is strictly convex, then for any $\bar{s}'_1$ satisfying Condition 7, $\bar{s}'_1 > \bar{s}_1$.*

This conclusion is startling: a doctor who tries to maximize the patient's expected utility does, in a sense, exactly the opposite of what he wants! A patient who wants to know the state of the world learns it less often than one who does not want to know it. This paradoxical outcome is the result of a subtle consideration affecting the first doctor's decision. In case she fails to inform the patient of his condition, she exposes him to *risk* stemming from what the second doctor will tell him. In particular, she runs the risk that the second doctor would be unable to come up with a diagnosis, and as a result he would not learn s at his second visit. He would then take this as bad news and feel bad about it. It is exactly this kind of risk that information-averse people do not like—essentially, risk-aversion over possible future beliefs is the definition of information-aversion. Thus, the first doctor would rather tell the information-averse patient the news herself, in order to shield him from the risk. Consequently, she reveals a larger set of diagnoses.[24]

Since u is strictly increasing, the equilibrium condition for the second doctor's cutoff diagnosis is the same for any u—Equation (5). From this condition, the second doctor's response to the first doctor's revealing more information is ambiguous.[25] However, even if the second doctor conveys less information to an information-averse patient, this is the result of her reaction to the first doctor's behavior, and—her behavior being only indirectly dependent on u—has nothing to do with her desire to comply with his informational preferences! The only doctor who responds to the patient's informational preferences is the first one—by going against them.

4.3 Information Disclosure with Treatment Choice

To complete the analysis of my model, I now extend the disclosure model to the case when the choice of treatment is non-trivial. I restrict attention to the one-period model. Thus, the communication game is now the following. The doctor observes s with probability α, and can either disclose it to the patient or send a message $m \in \mathcal{M}$ to him. After observing either d or m (whichever applies), he chooses $t \in \{0, 1\}$. I maintain the assumptions on the loss function introduced in Section 2.

[24] Since it is exactly the patient's information aversion that leads the doctor to reveal more information, a similar effect would be observed in many other models. In particular, the intuition is valid in any dynamic setting in which the party sending information today faces uncertainty about what information later informed parties would send.

However, the logic also indicates that uncertainty about what the second doctor will say is crucial for the result to hold. Consider what would happen if the second doctor knew the first one's diagnosis, or if the two were identical (e.g. the patient returns for more consultation to the same doctor). Although this situation may seem very similar to the previous one, there is a slight difference. Namely, if the first doctor chooses not to reveal s, she knows it can still be revealed in the second stage. Thus, even if she does not disclose s, she does not expose the patient's beliefs to (future) risk. Because of this, she has no reason to reveal more information to an information-averse patient (but the patient still learns more in the second stage). Interestingly, however, if the second doctor learns the first one's diagnosis, the first one reveals less information than in the model above. The proofs of these claims are omitted.

[25] How much the second doctor reveals depends on how "pessimistic" the patient is when his first visit ends with no disclosure; if he is more pessimistic, she reveals more information. A decrease in $\bar{s}_1'$, in turn, has two offsetting effects on the patient's level of pessimism when the first doctor does not offer a diagnosis. First, he attaches a larger probability to the possibility that the first doctor did not observe s. This makes him less pessimistic. On the other hand, *in case* the doctor did observe s, it is lower. This makes him more pessimistic. It is easy to see that either effect could dominate.

The following theorem summarizes the types of equilibria in this case.

Theorem 4. *If $\alpha = 1$, the doctor reveals all $s > 0$, and the correct treatment is always chosen. For $\alpha < 1$, the following type of perfect Bayesian equilibrium always exists. There is a $\bar{s} \in (0, 1)$ and an open set $N \subset [0, \bar{s}]$ such that i.) the doctor discloses all $s > \bar{s}$ and all $s \in N$; ii.) she does not disclose any s in the interior of $[0, \bar{s}] \setminus N$; iii.) she sends the same message whenever she does not disclose; and iv.) the patient chooses $t = 0$ for all $s \in N$.*

As the first statement in Theorem 4 indicates, for $\alpha = 1$ the doctor reveals all diagnoses, and the appropriate treatment is always chosen. This result contrasts sharply with that of Section 3, in which at most two distinct messages were sent, and often the wrong t was chosen.

For $\alpha < 1$, a new phenomenon emerges. In the new equilibrium, the diagnoses s that the doctor chooses to reveal are not necessarily better than the ones she chooses to hide. For example, it could be the case that she reveals good news, hides intermediate news, and also discloses the very worst news. Furthermore, these bad news are given only in states of the world where $t = 0$ is optimal. The doctor may prefer to be honest about bad news to induce the correct behavior. If she does not tell the patient how bad the situation is, he would remain too optimistic and choose $t = 1$ (with at least some probability). Thus, she "scares" him into realizing the gravity of his problem, and into doing the right thing. For example, if an athlete has a moderately serious knee injury, a doctor may not immediately want to disclose all details about how this is going to affect his future. But if the injury is so serious that he needs to take extra good care of his knee, she may be blunt and tell him all the bad news.[26]

As the proof of Theorem 4 indicates, there could also be a second type of simple perfect Bayesian equilibrium in the current version of the model. This equilibrium is a kind of "hybrid" between the above type and the equilibrium in Section 3. The doctor discloses all $s > \bar{s}$, and when she does not disclose s, she sends one of two possible messages depending on whether s is high or low. In addition, she might also scare the patient into choosing $t = 0$. This type of equilibrium requires the doctor to follow a mixed strategy when she does not observe s. If we insist on pure strategy

[26] In some instances, doctors seem to *exaggerate* patients' problems in order to get them to behave responsibly. See Footnote 32.

equilibria, only the first type of equilibrium survives.

5 Some Evidence

Medical and other research on doctor-patient relationships has devoted surprisingly little attention to how doctors transmit sensitive information to patients. The best existing evidence is Christakis (1999), who documents a pervasive "optimism bias" in the way doctors communicate life-threatening prognoses. In story after story, doctors recount their tendency to avoid communicating terrible news in a clear way, their strategy of firing "warning shots" and relaying even what they do say in stages, and their use of the inherent uncertainty in medicine as a vehicle for hope. They also emphasize their belief in the self-fulfilling prophecy (that optimistic prognoses can help physical outcomes). That this leads their patients to receive biased information (relative to what doctors really think) is evident from Table 1, reproduced from Christakis (1999). Strikingly, doctors are much more likely to communicate optimism about a given case when dealing with the patient than when talking to another doctor.[27]

Since there is very little evidence directly related to this paper's model, I have conducted a survey of practicing physicians on their experiences regarding doctor-patient communication. Although this is only preliminary evidence for the model, there is reason to believe that the survey results below *understate* the extent to which the theory's conclusions hold in reality. All the theoretical results in this paper derive from the doctor's incentive to manipulate the patient's beliefs; yet, in the United States, doctors are required to tell the truth. Thus, they may be reluctant to admit how much they are "tailoring" their information due to their patients' emotional needs.

Physicians were recruited on a voluntary basis, and were offered $50 for their participation in the half-hour phone survey. 73 physicians from California, Massachusetts, and New Jersey completed the final version of the survey.[28] A large portion of the survey questions was specifically

[27] Other existing evidence also indicates the common practice to "emphasize hope while remaining truthful about medical facts" (Ptacek and Eberhardt 1996, Quint 1977). Levine (1983) finds that doctors often delay conveying multiple sclerosis diagnoses to patients, for fear that their emotional reactions would aggravate the disease symptoms. As medical situations become more grim, information is less likely to be communicated (Fischbach 1983), and doctors describe various strategies to avoid discussing bad prognoses (Hopper and Fischbach 1989).

[28] The wording of some of the key questions in the initial pilot survey round was unclear, and the answers were

designed to test the foundations and predictions of the models in this paper. Table 2 summarizes answers to questions testing assumptions of the model. To start, doctors were asked in an open-ended question about the patient attributes that should be taken into account when communicating recommendations. 34 of the 45 doctors who gave specific answers—as opposed to generalities such as "each case is different"—mentioned the patient's emotional or mental health as one such attribute, and 25 of these physicians mentioned nothing else. This already suggest that doctors care about patients' emotions. But they were also asked specific questions probing whether they think emotions affect physical outcomes. If this is the case, even doctors who are intrinsically uninterested in their patients' feelings are in the end obligated to take them into account as well. Indeed, *all* 73 doctors said that fears affect physical health, and 72 of them believe in the placebo effect. When asked in an open-ended question about the ways in which fears affect physical health, doctors listed a variety of channels, including neurological and immunological mechanisms, depression, and insomnia. Among the more interesting specific examples were answers from an internist that fear increases palpitations and delays recovery, and an ophthalmologist's very practical concern that anxious patients cannot sit still for surgery. A few doctors also mentioned that fears affect compliance; and although not a priori obvious, most doctors seem to believe that fears impair rather than improve compliance (but see also further comments on compliance below).[29] More specifically, 71 doctors believe that fears affect patients' likelihood of monitoring themselves for symptoms, and 72 think that fears affect their likelihood of visiting their doctor. Once again, the majority of doctors (54 and 46, respectively) believe that fears tend to make it less likely that a patient checks for symptoms or visits a doctor, and a lot of the rest (13 and 16, respectively) think it depends on the situation.

Another important premise of my model is that patients can make reasonable inferences about their health from the doctor's information or recommendation. In the disclosure model, this amounts to being able to make inferences from *lack* of information. Thus, doctors were asked whether patients can figure out that there may be something seriously wrong with them if nothing

meaningless for the purposes of this paper. The wording was improved for the actual survey round, whose results are reported in this paper.

[29] Although compliance problems and other behavioral manifestations of fear are not modeled directly in this paper, they could clearly be one of the channels through which emotions act: if fear makes it less likely that a patient follows through a particular treatment, doctors do not want to suggest treatments that instill fear.

or little of their diagnosis is relayed to them. 70 of 73 physicians claim that patients are able to do so. In the model where the doctor cannot disclose her diagnosis, the important question is whether the patient can interpret the doctor's recommendation. The survey asked this for the specific example of sending the patient to a specialist.[30] The majority of respondents tended to think that patients can understand the informational content of this recommendation, but many more of them considered educated patients to be able to do so. Overall, support for the model's assumptions seems very strong.

More interesting, however, is the part of the survey that aimed to test predictions of the model. Results are summarized in Table 3. The first column of the table indicates the theorem to which the question applies, and the other four columns show the question and the doctors' answers. Regarding Theorem 1, 52 of 73 doctors claimed that their profession takes emotions into account when choosing treatments, with 17 doctors disagreeing and 4 providing no answer.[31] The 52 respondents who answered this question affirmatively were also asked to give examples of how patient anxiety might affect the chosen treatment. 40 doctors did so, and 38 of them gave examples consistent with the model. One internist mentioned her approach to dealing with a high, but not yet necessarily dangerous, blood sugar level. If the patient is an "emotional" type, she just recommends home exercises to watch sugar intake. However, if the patient is less fearful, aggressive nutrition education is in order. Another doctor mentioned that he is more likely to recommend "cutting back on" instead of "quitting" alcohol if the patient is afraid of cirrhosis. A neurologist said that tailoring the treatment to the patient's emotional vulnerability is common with neurological problems, since the patient's reaction is very important in those cases. Seven doctors mentioned that riskier but more effective treatments, or treatments with a lot of "psychological baggage,"

[30] The question was the following: "Imagine that a patient comes in with a symptom she considers suspicious. If you think the symptom is not serious, you might send her home. If it is serious, you might refer her to a specialist. If you refer her to a specialist, will she understand that she is more likely to be sick?"

[31] For this key question in the survey, it was necessary to give doctors an example, and to emphasize that the question asks specifically about the suggested treatment. In pretesting, respondents often misunderstood the question otherwise. The final wording of the question is the following: "Suppose a smoker goes to a doctor with lung problems. If she is told that she needs to quit smoking immediately, she might get frightened and think that she has a very serious problem, while if she is told that she should simply try to start cutting down on her smoking, she might take the news more calmly. With this example in mind, do you think that when prescribing treatments, doctors take into consideration the ways in which their patient will interpret their recommendations? Let me emphasize that I am talking about the actual **treatment**, not how the treatment is presented to the patient."

are not given to more anxious patients. Other examples include cigarette smoking cessation and addiction (13 doctors); weight loss, diabetes, anorexia, bulimia, and other diet related conditions (14); exercise (4); glaucoma (2); dangers of sun, contact lens care, mental health, and even cancer treatments (1 each). Most of these are exactly the kinds of conditions to which the model of Section 3 applies: optimal treatments are informative about the patient's health, and the loss function is not very steep.

It is worthwhile to emphasize what these doctors are saying. They are asserting that they suggest *less effective treatments* to patients who are more emotional. While completely natural in the context of this paper, this assertion runs counter to standard economic models. It also may explain why, despite the high usage and quality of intensive medical services (Cutler 2002), prevention and chronic disease management receives comparatively little attention in the United States (e.g. McBride et al 1998, Legorreta et al 1998). As the model predicts, this is often because doctors fail to counsel patients on the right course of action.[32]

Of the 17 doctors who did not believe doctors adjust their treatments due to patients' emotions, 8 suggested that the reason for this is the primacy of physical considerations over emotional ones, as in Lemma 3. Finally, the two doctors who gave examples that are inconsistent with the model also emphasized that they would recommend less severe treatments to more emotional types, but for a completely different, compliance related reason: they said that less fearful types tend to be more overconfident, and, as a consequence, it takes a more drastic recommendation for them to change at all. Despite this, the support for the predictions of my first model is strong.

There was much more disagreement between physicians in the responses to questions aimed mostly at testing Theorems 2 and 3. Although most doctors agreed that negative information is communicated differently than positive information, only 37 of these thought that bad news are communicated in a dampened way, and 32 thought that they are communicated with more delay. Although the support for the disclosure model is thus not unanimous, I can use the variation in

[32] Interestingly, when *explaining* their recommendations, doctors often *exaggerate* the patient's problems. For example, they may tell him that he could die if he does not take the prescribed medicine. While this has the flavor of Theorem 4's prediction that the doctor might shock the patient into behaving correctly, there are other reasons—such as denial or self-control problems on his part—to do it as well. These latter considerations are not captured in this paper.

doctors' answers to assess the determinants of their diverse opinions, as well as to test a more specific prediction. Namely, whether or not the doctor can certifiably reveal her diagnosis, the model predicts that she is less likely to dampen or delay relaying negative information if that is costly from a medical health point of view. Thus, doctors who work in fields where problems are generally more acute should be less likely to answer these questions affirmatively.

To test this prediction, I define a binary variable that takes the value of 1 if the physician asserted either that negative information is communicated with more delay, or in a more dampened way, than positive information. (The key coefficients have the same sign if I use each of these variables separately, but they are smaller and have lower significance.) Based on their fields, I separate doctors into two groups, depending on whether they tend to deal with less or more acute problems. Mostly general internists, dermatologists, gynecologists, ophthalmologists, and psychiatrists are in the first group, and the second group consists largely of cardiologists, endocrinologists, neurologists, and oncologists. Since this division is somewhat subjective, I also perform the analysis splitting doctors into general internists and specialists.

The results of probit analyses of doctors' answers are presented in Table 4. The top panel presents results with the broader categorization of "less acute" fields, and the bottom panel with the general internist dummy. Column I gives the results of a probit regression using the above field dummy as the only regressor. In strong support of the model, physicians in the first group of the fields above are significantly more likely to say that negative information is communicated in a dampened way or with delay. As controls are added to the regression in Columns II through V, the coefficient on the field dummy only increases, and becomes more significant. A similar pattern can be observed in the bottom panel.

Although not part of the test of the model, it is interesting to note how the doctor's gender and experience affect the likelihood of dampening or delaying negative information. Surprisingly, specification IV seems to indicate that male physicians are more likely to soften information. It is perhaps less surprising that more experienced doctors more often say they dampen or delay information, since standards about truth-telling have changed considerably over the last 30-40 years. However, if we break up the effect of experience into that of males and females separately, it is clear that this result is driven mostly by more experienced females. Combining the two results, this indicates that more experienced females and less experienced males are more likely to dampen or delay negative information![33]

While it is a question whether some of these results hold up in larger and more complete datasets, in sum they strongly suggest that the effects discussed in this paper are real and important. In addition, as emphasized above, relying on doctors' self-reports may only provide a lower bound on the extent of these phenomenona. Of course, future work should expand the survey, explore this problem from both doctors' and patients' points of view, and consider behavioral evidence in addition to survey data.

6 Conclusion

It is hardly a matter of dispute that patients facing a medical situation experience emotions regarding their future health, and that doctors care about these feelings when making decisions. I argue in this paper that these emotions, even if they depend *exclusively* on anticipated physical outcomes, are inherently different from such outcomes: they motivate the doctor to (often counter-productively) manage the patient's feelings. This leads to a number of predictions that run counter to economic analysis with standard payoffs: the best-intentioned doctor, in order to reassure her patient, might too often not recommend a painful lifestyle change; it is more difficult to give good recommendations when feelings are a larger part of the patient's welfare; and—unless it is necessary to shock him into doing the right thing—bad news are held back from the patient for a while, but can come out eventually. In addition, this framework makes novel predictions as well. Surprisingly, ignorant patients might get better care than better-educated ones, and information-averse patients often receive *more* information in equilibrium than information-loving ones.

The results in this paper suggest at least three avenues for future research. First and foremost, the communication problems faced by the doctor and the patient suggest a need for an analysis of mechanisms that can alleviate them.[34] In my model, the doctor is assumed to arrive at her diagnosis

[33] Using the point estimates in the regression with the acute fields dummy, the crossover occurs around 20.5 years of experience. Using the point estimates from the other regression, it occurs at 15.1 years of experience.

[34] See Caplin and Eliaz (2002) for a public policy design problem when the population is motivated by fear of bad

instantaneously, and to have only one chance to communicate with the patient. In reality, there is more flexibility in how doctors and patients communicate, and in how a final diagnosis is reached. Future work can offer guidelines for the design of this process. A basic question is whether partial information should be communicated before the doctor identifies the precise problem. In some cases at least, the answer seems to be yes: giving early information alleviates later communication problems, and the doctor may not dare to mislead the patient based on partial knowledge. For example, a family practitioner, who can only make a broad diagnosis of a complicated condition, does not want to send the patient home when her preliminary diagnosis calls for sending him to a specialist. But once the family practitioner has, by referring him to another doctor, broken some bad news to the patient, the specialist is less motivated to distort her own recommendation. This advantage of relaying early *specific* information is a counterpart to Lemma 2's result that it is better for the patient not to know *general* properties of his medical condition.

Second, relatedly, it would be interesting to study how various forms of education about medical issues affect fears, especially since—according to many doctors in the survey—more educated patients tend to be more fearful. Sorting out these issues may be important in understanding medical outcomes in the face of educational improvements and the vast increase in the availability of medical information due to the internet.

Finally, one more limitation of the model deserves comment. Namely, the patient cannot choose his doctor. Competition between doctors is unlikely to reliably eliminate the basic problems in this paper, but it would be exciting to identify conditions under which they are mitigated or exacerbated.

news in HIV testing.

	Physician's statement to colleague			
	pessimistic	neither	optimistic	Total
Physician's statement to patient, %				
pessimistic	36.5	3.0	0.5	39.9
neither	16.8	2.5	0	19.2
optimistic	15.3	14.3	11.3	40.9
Total	68.5	19.7	11.8	100

203 internists received a vignette describing a sixty-six-year-old man with chronic obstructive pulmonary disease and pneumonia necessitating ICU admission. Subjects were queried regarding how much prognostic optimism or pessimism they would communicate to the patient and to a colleague. **(reproduced from Christakis (1999), page 228)**

Table 1: Prognostic Optimism Communicated to Patients and Colleagues

Question	# Yes	# No	# No Answer
Mentioned emotional state of patient as determinant of how recommendations are communicated	34	11	28
Does fear of illness affect physical health?	73	0	0
Do you believe in the placebo effect?	72	1	0
Does fear of illness affect the patient's likelihood of checking for symptoms?	71	1	1
Does fear of illness affect the patient's likelihood of visiting a doctor?	72	1	0
Do patients interpret the lack of information negatively?	70	3	0
Can patients draw inferences from the doctor's recommendation?			
Patient educated in medicine	65	8	0
Patient educated, but not in medicine	56	17	0
Patient uneducated	38	35	0

(Full survey available from the author.)

Table 2: Summary of Survey Results On Assumptions of the Model

Relevant Result	Question	# Yes	# No	# No Answer
Theorem 1	Do doctors adjust their recommended treatments due to the patient's emotions?	52	17	4
Theorem 1, Corollary 1	Of 52 "Yes", provided example consistent with the model	38	2	12
Lemma 3	Of 17 "No", mentioned as reason that physical problems outweigh interpretational considerations	8	5	4
—	Is negative health information communicated differently than positive information?	70	3	0
	Of 70 "Yes", mentioned that negative information is often communicated ...			
—	... by a different person	37	33	0
Theorem 2	... in a more dampened way	34	36	0
—	... only when medically necessary	5	65	0
Theorem 3	... with more delay	32	38	0
Theorems 2,3	Of 70 "Yes", considers emotional state of patient as prime determinant of how bad news are communicated	39	31	0

NOTE: A response was coded as "No Answer" if the doctor either failed to provide a clear answer, or provided one that clearly reflected a lack of understanding or attention to the question.
(Full survey available from the author.)

Table 3: Summary of Survey Results on Predictions of the Model

Dependent Variable: Physician Dampens or Delays Negative Information

	I	II	III	IV	V	VI
acute disease field	0.251*	0.414**	0.424**			
	(0.117)	(0.129)	(0.128)			
internal medicine				0.221*	0.290*	0.338**
				(0.103)	(0.100)	(0.100)
male		0.280*	0.575*		0.190	0.598*
		(0.136)	(0.217)		(0.127)	(0.220)
experience (in years)		0.017*			0.016*	
		(0.007)			(0.007)	
male experience			0.012			0.010
			(0.008)			(0.008)
female experience			0.039*			0.048*
			(0.018)			(0.020)
Pseudo R^2	0.051	0.191	0.215	0.045	0.154	0.193

⋄ * and ** denote significance at 5% and 1%, respectively.
⋄⋄ Marginal probit coefficients are shown.
⋄ ⋄ ⋄ Field dummy takes the value of 1 if doctor's field is dermatology, general internal medicine, gynecology, ophthalmology, orthopedics, psychiatry, rheumatology, or urology; it takes the value of 0 if doctor's field is cardiology, cosmetic surgery, endocrinology, gastroenterology, hematology, neurology, or oncology.

Table 4: Determinants of Dampening or Delaying the Communication of Bad News

A Definition of Perfect Bayesian Equilibrium

To define perfect Bayesian equilibrium formally, let $\sigma_d(s,m)$ be the probability that the doctor sends message m when her diagnosis is s, $\sigma_p(m,t)$ the patient's probability of choosing treatment t after message m, and $\mu(m)$ his beliefs about s after message m.

Definition 1. $\sigma_d(\cdot,\cdot)$, $\sigma_p(\cdot,\cdot)$, and $\mu(\cdot)$ constitute a perfect Bayesian equilibrium if

1. **Doctor optimization**—For all $s \in [0,1]$, if $\sigma_d(s,m) > 0$, then

$$m \in \text{argmax}_{m' \in \mathcal{M}}\, \sigma_p(m',0)U(\mu(m'),s,0) + \sigma_p(m',1)U(\mu(m'),s,1).$$

2. **Patient optimization**—For all $m \in \mathcal{M}$, if $\sigma_p(m,t) > 0$, then

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

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

A perfect Bayesian equilibrium is simple if for all $m, m' \in \mathcal{M}$,

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

B Proofs

Proof of Lemma 1. We first show by contradiction that in any simple perfect Bayesian equilibrium, no two messages that are sent with positive probability lead the patient to use the same strategy. If there were two such messages, they would have to induce the same anticipatory utility—otherwise the doctor would never choose the one that induces lower anticipatory utility. But in that case, the equilibrium would not be simple, a contradiction.

Therefore, we can order the doctor's messages by the probability that the patient will choose $t = 1$ after the message. Clearly, since $L(s)$ is strictly decreasing, the doctor's message is increasing with s in this order, and the doctor adopts a pure strategy for almost all s. Now we distinguish two cases, depending on whether there is a message that leads the patient to mix between the two treatments.

Suppose there is a message m sent by the doctor with positive probability that leads the patient to mix. Then, the set of diagnoses s where the doctor sends m must include some $s > s^*$; otherwise the patient could not be indifferent between $t = 0$ and $t = 1$. Then, there is no message that leads the patient to choose $t = 1$ with a higher probability: if there was such a message, the doctor would prefer to send it for any $s > s^*$ for both anxiety and health reasons. This also implies that the doctor sends m for any $s > s^*$. Therefore, any other message leads the patient to choose $t = 0$ with probability 1. And, as noted above, in a simple perfect Bayesian equilibrium there can only be one such message.

Now suppose that the patient follows a pure strategy after all messages. Then, again from the above, in a simple perfect Bayesian equilibrium, there can only be two messages that are sent with positive probability. □

Proof of Theorem 1. Suppose the two messages are m_0 and m_1, and the patient chooses $t = i$ with probability 1 after m_i. First, since $L(s)$ is decreasing, it is clear that the doctor follows a cutoff strategy (sending m_1 for diagnoses above a given cutoff), no matter what the patient's beliefs. Call the cutoff value of the diagnosis s^c.

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

$$\max_{t\in\{0,1\}} \int h(s,t)d\mu(m_1)(s) = \int_{s^c}^{1} sdF(s) > \int_{0}^{s^c} sdF(s) \geq \max_{t\in\{0,1\}} \int h(s,t)d\mu(m_0)(s).$$

Therefore, for $s = s^*$, the doctor, 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 diagnoses slightly below s^c. This contradicts $s^c \geq s^*$. □

Proof of Corollary 1. Let $s^c(w)$ be the cutoff in the best perfect Bayesian equilibrium for a given w. I prove that $s^c(w)$ is strictly decreasing.

On the interval $[0, s^*]$, consider the curves $L(s)$ and

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

The cutoff in the best perfect Bayesian equilibrium of our game, $s^c(w)$, is given by the highest intersection of the curves $wA(s)$ and $L(s)$. Since $A(s)$ and $L(s)$ is continuous and $A(s^*) > L(s^*) = 0$, we know that $wA(s) > L(s)$ 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 $wA(s^c(w)) = L(s^c(w)) > 0$, there is some $\epsilon > 0$ such that $w'A(s) > L(s)$ for any $s > s^c(w) - \epsilon$. This means that $s^c(w') \leq s^c(w) - \epsilon$. □

Proof of Lemma 2. Take any two $m_0, m_1 \in \mathcal{M}$ with $m_0 \neq m_1$. Let $\sigma_d(s, m_0) = 1$ for $s < s^*$, and $\sigma_d(s, m_1) = 1$ for $s \geq s^*$. That is, the doctor sends message m_0 when $t = 0$ is appropriate, and m_1 when $t = 1$ is appropriate. Let the patient's strategy be to "follow" this advice: $\sigma_p(m_i, i) = 1$ for $i = 1, 2$, with any arbitrary strategy for the rest of the messages. It is easy to check that there are beliefs that make these strategies part of a perfect Bayesian equilibrium. Namely, for any $x \in [0, 1]$, let $\mu(m_0)(x) = \frac{\min\{F(x), F(s^*)\}}{F(s^*)}$ and $\mu(m_1)(x) = \frac{\max\{0, F(x) - F(s^*)\}}{1 - F(s^*)}$, and for any other message, beliefs attach probability 1 to the worst outcome. □

Proof of Lemma 3. We prove that for any $\epsilon > 0$, there is a $\delta > 0$ such that if $L^{-1}(w) > s^* - \delta$, in any informative perfect Bayesian equilibrium the patient's expected health is within ϵ of the optimum $E[s]$.

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

$$w\left[E(s - l(s, 1)|s \geq s^c) - E(s|s \leq s^c)\right] \leq w. \tag{8}$$

Therefore, if $L^{-1}(w) > s^* - \delta$, any informative equilibrium cutoff s^c satisfies $s^* - \delta < s^c < s^*$. Furthermore, $l(s, 1) < w$ for any $s \in [s^c, s^*]$. Therefore, the loss in expected expected relative to the optimum is bounded from above by $w\left(F(s^*) - F(s^c)\right)$. For a sufficiently small δ, this is clearly less than ϵ. □

Proof of Lemma 4. Suppose m_1 and m_2 are two messages sent by the doctor when she does not disclose s. Then, they induce the same anticipatory utility, since otherwise the doctor would not want to send the one that induces the lower one. By assumption, the patient also follows the same strategy after the two messages. Thus, if we are in a simple perfect Bayesian equilibrium, m_1 and m_2 must be the same message. □

Proof of Theorem 2. We look for the equilibrium value of v, the anticipatory utility of the patient when the doctor does not reveal s. Clearly, for any given v, the doctor reveals s if and only if $s > v$. 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 agent's anticipatory utility when the doctor does not disclose s, and he believes that she discloses diagnoses $s > v$. Clearly, we have an equilibrium 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 equilibrium. Now

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

Thus, whenever $V(v) = v$, $V'(v) = 0$. This proves that the equilibrium value of v, v^*, is unique. $\bar{s} = v^*$ satisfies the requirements of the theorem. □

Lemma 5. *In any simple perfect Bayesian equilibrium with two doctors in a row communicating with the patient, both doctors send just one message each when they do not reveal s.*

Proof. For $\alpha = 1$, it is obvious (using essentially the same proof as in Theorem 2) that both doctors reveal all diagnoses greater than zero. We continue with the case $\alpha < 1$, where it is obvious that both doctors disclose the diagnosis with probability less than 1.

By the proof of Theorem 2, after any message of doctor 1, the continuation equilibrium in a simple perfect Bayesian equilibrium has the following properties: doctor 2 follows a cutoff strategy, and she always sends the same message when she does not disclose s. We prove that all messages that are sent with positive probability by doctor 1 lead to the same stage-1 anticipatory utility, and they lead doctor 2 to adapt the same cutoff. This is clearly sufficient, since it implies that in a simple perfect Bayesian equilibrium doctor 1 sends the same message whenever she does not disclose s.

Let m be the message that leads doctor 2 to adopt the highest cutoff, and let m' be the message that leads her to adopt the lowest one. Let these cutoffs be $\tilde{s}$ and $\tilde{s}'$, respectively, and suppose the messages m and m' lead to stage 1 anticipatory utilities of u and u', respectively. Notice that, by the equilibrium condition for doctor 2's cutoff, the induced stage 2 feelings in the two continuation equilibria when doctor 2 does not reveal s are $\tilde{s}$ and $\tilde{s}'$, respectively.

Think of the case $\tilde{s} = \tilde{s}'$ first. Then, it must be the case that $u = u'$. If, say, $u > u'$, the doctor would never want to send message m': m leads to higher anticipatory utility in stage 1, and the same expected anticipatory utility in stage 2. Thus, we are done in this case.

To complete the proof, we prove by contradiction it is impossibe to have $\tilde{s} > \tilde{s}'$. If that was the case, then whatever is the diagnosis s, the patient's expected anticipatory utility at stage 2, from the point of view of doctor 1, is higher after message m than after message m'. To see this, note that the respective utilities are $\tilde{s}$ and $\tilde{s}'$ for $s < \tilde{s}'$ (since the second doctor will not disclose s even if she knows it), they are $\tilde{s}$ and $\alpha s + (1-\alpha)\tilde{s}'$ for $\tilde{s}' \leq s < \tilde{s}$, and $\alpha s + (1-\alpha)\tilde{s}$ and $\alpha s + (1-\alpha)\tilde{s}'$ for $s \geq \tilde{s}$. Thus, in order for doctor 1 to be willing to send the message m', it has to be the case that $u' > u$.

It is easy to prove that in order for doctor 1 to be willing to send m for any diagnosis s, she must send m for all $s < \tilde{s}'$. For the doctor to be willing to send m, it must be prefered over any $m'' \in \mathcal{M}$. According to the calculation in the previous paragraph, the greatest possible benefit from sending m, which leads to lower utility in stage 1, but higher expected utility in stage 2, is when $s < \tilde{s}'$ (since in that case doctor 2 does not disclose s after any message). Furthermore, it cannot be the case that she is indifferent between m and another message in this range. To see this, note that when doctor 1 does not observe s, she has to be willing to send all messages that are sent with positive probability in equilibrium. From the point of view of doctor 1 who does not know s, the probability that doctor 2 reveals s is positive. However, this probability is zero if she knows s and $s < \tilde{s}'$, so in that range sending m is strictly preferred.

Now we know that $\tilde{s}'$ is the cutoff in the continuation equilibrium after m'. But since the message m is sent whenever $s < \tilde{s}'$, $\mu(m)$ has more weight below $\tilde{s}'$ than $\mu(m')$. But this implies that the equilibrium cutoff following m is lower than $\tilde{s}'$, a contradiction. □

Lemma 6. *Suppose the patient sequentially visits two doctors, and that each doctor sends just one message when she does not reveal* s. *In any perfect Bayesian equilibrium, both doctors follow a cutoff strategy: there are* $\bar{s}_1, \bar{s}_2 \in [0, B]$ *such that doctor* i *reveals* s *whenever* $s \geq \bar{s}_i$.

Proof. It follows from Lemma 5 that doctor 2 always follows the same cutoff strategy. Given this, the first doctor also follows a cutoff strategy: she reveals s when it is greater than the patient's expected feeling that would be realized otherwise. □

Proof of Theorem 3. We first prove that in any perfect Bayesian equilibrium $\bar{s}_1 \geq \bar{s}_2$. By contradiction, assume that $\bar{s}_1 < \bar{s}_2$. Then, $\bar{s}_1$ and $\bar{s}_2$ must satisfy

$$\begin{aligned} \bar{s}_2 &= \frac{(1-\alpha)^2 E[s] + \alpha F(\bar{s}_1) E[s|s<\bar{s}_1] + \alpha(1-\alpha) F(\bar{s}_2) E[s|s<\bar{s}_2]}{(1-\alpha)^2 + \alpha F(\bar{s}_1) + \alpha(1-\alpha) F(s<\bar{s}_2)} \\ 2\bar{s}_1 &= \frac{(1-\alpha)E[s] + \alpha F(\bar{s}_1) E[s|s<\bar{s}_1]}{(1-\alpha)+\alpha F(\bar{s}_1)} + \bar{s}_2. \end{aligned} \tag{9}$$

The first of these is the second doctor's cutoff condition—when diagnosing $s = \bar{s}_2$, she must be indifferent between revealing this and remaining silent. The right-hand side of this expression is the patient's anticipatory utility given that neither doctor has revealed any information, and given their strategies. The second expression is the first doctor's cutoff condition. If she reveals $\bar{s}_1$, in both stages that will be the patient's anticipatory utility. If she doesn't reveal $\bar{s}_1$, since $\bar{s}_1 < \bar{s}_2$ she knows that the second doctor will not reveal it, either, so in the second stage the patient will realize the anticipation after both doctors have revealed nothing. By the second doctor's indifference condition, this is equal to $\bar{s}_2$—the second term on the right-hand side. The first term is the patient's anticipatory utility in the first stage after not hearing a diagnosis, given the first doctor's strategy.

Now since $\bar{s}_1 < \bar{s}_2$, we have

$$\begin{aligned} &\underbrace{\frac{(1-\alpha)^2 E[s] + \alpha F(\bar{s}_1) E[s|s<\bar{s}_1] + \alpha(1-\alpha) F(\bar{s}_2) E[s|s<\bar{s}_2]}{(1-\alpha)^2 + \alpha F(\bar{s}_1) + \alpha(1-\alpha) F(\bar{s}_2)}}_{\text{I}} \\ &= \bar{s}_2 > \bar{s}_1 > \underbrace{\frac{(1-\alpha)E[s] + \alpha F(\bar{s}_1) E[s|s<\bar{s}_1]}{(1-\alpha)+\alpha F(\bar{s}_1)}}_{\text{II}}, \end{aligned} \tag{10}$$

where the last inequality follows from the first doctor's cutoff condition. To arrive at a contradiction, we prove that the above is impossible, that we cannot have $I > II$. Intuitively, the patient cannot feel better after both doctors have failed to give her a diagnosis than after only one has, since a doctor's failure to inform her might indicate that she is quite sick. Formally, rewrite I as

$$\frac{(1-\alpha)(1-\alpha+\alpha F(\bar{s}_2)) \left(\frac{(1-\alpha)E[s]+\alpha F(\bar{s}_2)E[s|s<\bar{s}_2]}{(1-\alpha)+\alpha F(\bar{s}_2)} \right) + \alpha F(\bar{s}_1) E[s|s<\bar{s}_1]}{(1-\alpha)(1-\alpha+\alpha F(\bar{s}_2)) + \alpha F(\bar{s}_1)}. \tag{11}$$

This form facilitates the comparison between I and II. Both expressions feature the weighted average of $E[s|s < \bar{s}_1]$ and a larger number, and the weight on $E[s|s < \bar{s}_1]$ is the same. Now notice that in expression 11, the value $E[s|s < \bar{s}_1]$ is being averaged with is smaller (than $E[s]$, that in expression II) and the weight placed on it is also smaller (than $1 - \alpha$.) Therefore, we have $I < II$, completing the proof that $\bar{s}_1 \geq \bar{s}_2$.

Now it is easy to write the expression that $\bar{s}_1$ and $\bar{s}_2$ must satisfy. The second doctor's cutoff condition is similar to before. There is a slight difference in the first doctor's cutoff condition. If she chooses not to inform the patient, then she does not know what will happen in the second period. Since $\bar{s}_1 \geq \bar{s}_2$, if the second doctor learns s, she will communicate it, giving the patient an anticipatory utility of $\bar{s}_1$. However, if she is unable to diagnose the patient, she cannot reveal anything, so the patient will get $\bar{s}_2$. Thus the second two terms in the first doctor's cutoff condition.

It is easy to check that we cannot have $\bar{s}_1 = \bar{s}_2$. Thus, in any equilibrium, $\bar{s}_1 > \bar{s}_2$.

Finally, we prove that the equilibrium is unique. By Theorem 2, for any value of s_1, there is a unique $\bar{s}_2(s_1)$ that satisfies the first of Expression 6 (with $\bar{s}_1$ replaced by s_1). Furthermore, $\bar{s}_2(s_1)$ is differentiable, and, just as in the proof of Theorem 2, it is easy to show that

$$\bar{s}_2'(s_1) = \frac{\alpha(1-\alpha)f(s_1)(s_1 - \bar{s}_2(s_1))}{(1-\alpha)^2 + \alpha F(\bar{s}_2(s_1)) + \alpha(1-\alpha)F(s_1)}. \tag{12}$$

Let

$$V(s_1) = \frac{(1-\alpha)E[s] + \alpha F(s_1)E[s|s < s_1]}{(1-\alpha) + \alpha F(s_1)}.$$

Once again, as in the proof of Theorem 2,

$$V'(s_1) = \frac{\alpha f(s_1)(s_1 - V(s_1))}{(1-\alpha) + \alpha F(s_1)}. \tag{13}$$

Now, equilibrium is defined by

$$(2-\alpha)s_1 = V(s_1) + (1-\alpha)\bar{s}_2(s_1). \tag{14}$$

For $s_1 = 0$, the right-hand side is greater, for $s_1 = 1$, the opposite is the case. By continuity, there is an equilibrium. We have already established that in this case $s_1 - \bar{s}_2(s_1) > 0$. Furthermore, whenever the two sides of Equation 14 are equal, $s_1 - V(s_1) = (1-\alpha)(\bar{s}_2(s_1) - s_1)$. Using this and

Expressions 12 and 13, the derivative of the right-hand side of 14 at an equilibrium is

$$(1-\alpha)\left(\frac{\alpha f(s_1)(\bar{s}_2(s_1)-s_1)}{(1-\alpha)+\alpha F(s_1)}+\frac{\alpha(1-\alpha)f(s_1)(s_1-\bar{s}_2(s_1))}{(1-\alpha)^2+\alpha F(\bar{s}_2(s_1))+\alpha(1-\alpha)F(s_1)}\right).$$

Since $s_1-\bar{s}_2(s_1)>0$, the above derivative is negative. This establishes that equilibrium is unique. $\square$

Proof of Corollary 2. We prove for a concave u. The proof for a convex u is symmetric. We can define $\bar{s}_2(s_1)$ as in the proof of Theorem 3. Since $\bar{s}_2(s_1)$ does not depend on the shape of u, an equilibrium in this case is defined by

$$(2-\alpha)u(s_1)=u(V(s_1))+(1-\alpha)u(\bar{s}_2(s_1)). \tag{15}$$

By Jensen's inequality, whenever the right-hand side of equation 15 is smaller than the left-hand side for a linear u, it is smaller for a concave u. Thus, any s_1 that is an equilibrium for a concave u is smaller than the equilibrium value of s_1 for a linear u. Furthermore, since u is strictly concave, we obviously cannot have $\bar{s}_1=\bar{s}'_1$. $\square$

Proof of Theorem 4. The proof for the $\alpha=1$ is almost identical to that of Theorem 2. Assume by contradiction that there is a set N of possible diagnoses s with measure greater than zero for which the doctor does not disclose her information. Then, there is a subset of non-zero measure of values s for which $s>E[s'-l(s',t)|s'\in N]$, where t is any treatment chosen when the doctor does not disclose s. So, by revealing s in those states, the doctor can make the patient feel better, a contradiction.

Now consider $\alpha<1$. Notice first that by the same argument as above, any message in $\mathcal{M}$ that is sent with probability zero when the doctor is ignorant of s is sent with probability zero overall. Thus, the ignorant doctor is indifferent between sending all messages that are sent with positive probability in equilibrium. Call this set of messages $\mathcal{M}_0$.

Next, we prove that in a simple perfect Bayesian equilibrium, $\mathcal{M}_0$ has at most two elements. Assume that $E[l(s,1)]>E[l(s,0)]$ (that is, when the doctor does not make a diagnosis, $t=0$ is optimal); a similar proof works in the opposite case. Each message in $\mathcal{M}_0$ induces some anticipatory utility in the patient, and leads to him choosing $t=1$ with a given probability. Since the doctor prefers $t=0$ when she does not make a diagnosis, for her to be indifferent between the messages in $\mathcal{M}_0$, it has to be the case that if $m \in \mathcal{M}_0$ leads to a higher probability of choosing $t = 1$ than $m' \in \mathcal{M}_0$ ($\sigma_p(m, 1) > \sigma_p(m', 1)$), then it also induces a higher anticipatory utility. As a consequence, whenever the doctor does make a diagnosis, and $s > s^*$, she prefers to send the message $m \in \mathcal{M}_0$ that leads to the highest probability of choosing $t = 1$ (since this also leads to the best feeling). Thus, when she sends any other message, the patient knows that $s < s^*$, and chooses $t = 0$ with probability one. This implies that any message other than m leads to the same anticipatory utility in the patient—otherwise the doctor would just want to send the one that leads to the highest anticipatory utility. Thus, in a simple perfect Bayesian equilibrium, the doctor sends just at most one message other than m with positive probability.

Now we have two cases, depending on whether $\mathcal{M}_0$ has one or two elements (it clearly cannot have zero elements for $\alpha < 1$). First assume it has one element m_1. Let $\bar{s}$ be the greatest lower bound of all $s \in (0, 1)$ such that the doctor discloses all diagnoses in $(s, 1)$. It is easy to prove by contradiction that for any $s < \bar{s}$ that the doctor discloses, $s < s^*$ (and, consequently, the disclosure leads the agent to choose $t = 0$): if she dislosed some $s^* \leq s < \bar{s}$, she would also want to disclose any diagnosis on the interval $(s, 1)$, since disclosing the larger diagnosis is more justified on both anticipation and physical health grounds. But this would contradict the construction of $\bar{s}$. By the same argument, if $\sigma_p(m_1, 1) = 0$, the doctor discloses no $s < \bar{s}$.

Let u be the anticipatory utility induced by the message m_1: $u = \max_{t\in\{0,1\}} w \int (s-l(s, t))d\mu(m_1)(s)$. Then, the doctor discloses the following set of diagnoses below $\bar{s}$: $N = \{s < s^* | u - ws < \sigma_p(m_1, 1)l(s, 1)\}$. N is clearly open and satisfies the requirements of the theorem. Thus, in this case the equilibrium is of the type described in the theorem.

Now suppose that $\mathcal{M}_0$ has two elements m_1 and m_2, $\sigma_p(m_1, 1) < \sigma_p(m_2, 1)$ (that is, m_1 leads to the patient to choose $t = 1$ with lower probability.) Again, assume that $E[l(s, 1)] > E[l(s, 0)]$; the proof for the opposite case is similar. Then, by the same logic as above, m_1 leads patient to choose $t = 0$ with probability 0 ($\sigma_p(m_1, 1) = 0$), and m_2 leads him to choose $t = 1$ with positive probability. As before, let $\bar{s}$ be the greatest lower bound of all $s \in (0, 1)$ such that the doctor discloses all diagnoses in $(s, 1)$. It is easy to see that any diagnosis $s < \bar{s}$ that the doctor discloses satisfies $s < s^*$. It is also obvious that the doctor follows a cutoff strategy if she does not disclose

s. Thus, in this case the equilibrium has the following properties:

> There are $\bar{s} \in (0,1)$, $s^c \in (0, \bar{s})$, and an open set $N \subset [0, \bar{s}]$ such that i.) the doctor discloses all $s > \bar{s}$ and all $s \in N$; ii.) she does not disclose any s in the interior of $[0, \bar{s}] \setminus N$; iii.) when she observes s and does not disclose it, she sends a message m_1 if $s \in [0, s^c)$ and a message $m_2 \neq m_1$ if $s \in [s^c, \bar{s})$; iv.) the patient chooses $t = 0$ for all $s \in N$; v.) either $N \subset [0, s^c]$ or $N \subset [s^c, \bar{s}]$; and vi.) the doctor randomizes between m_1 and m_2 when she does not observe s.

Finally, we prove that the first type of equilibrium always exists. We look for an equilibrium pair v, p, where v is the patient's anticipatory utility when the doctor does not disclose s ($v = \max_{t \in \{0,1\}} w \int (s - l(s,t)) d\mu(m_1)(s)$), and p is the probability that he chooses $t = 1$ in the same situation ($p = \sigma_p(m_1, 1)$). (Since the doctor always sends the same message when she does not disclose s, these numbers are well-defined.) Let

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

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

$$V(v,p) = \frac{\alpha E[s - pl(s,1) - (1-p)l(s,0) | s \in N(v,p)] + (1-\alpha) E[s - pl(s,1) - (1-p)l(s,0)]}{\alpha Prob(s \in N(v,p)) + (1-\alpha)},$$

which is the patient's expected anticipatory utility if she expects the doctor not to disclose diagnoses in $N(v,p)$. Also, define the correspondence $T : \mathbb{R} \times [0,1] \rightrightarrows \{0,1\}$ by

$$0 \in T(v,p) \quad \text{iff} \quad \frac{\alpha E[l(s,1) | s \in N(v,p)] + (1-\alpha) E[l(s,1)]}{\alpha Prob(s \in N(v,p)) + (1-\alpha)} \geq \frac{\alpha E[l(s,0) | s \in N(v,p)] + (1-\alpha) E[l(s,0)]}{\alpha Prob(s \in N(v,p)) + (1-\alpha)}, \text{ and}$$

$$1 \in T(v,p) \quad \text{iff} \quad \frac{\alpha E[l(s,1) | s \in N(v,p)] + (1-\alpha) E[l(s,1)]}{\alpha Prob(s \in N(v,p)) + (1-\alpha)} \leq \frac{\alpha E[l(s,0) | s \in N(v,p)] + (1-\alpha) E[l(s,0)]}{\alpha Prob(s \in N(v,p)) + (1-\alpha)}.$$

That is, $T(v,p)$ is the set of treatments the patient is willing to choose when she expects the doctor not to disclose the diagnoses in $N(v,p)$. Clearly, v, p constitutes an equilibrium 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 $0 < p < 1$.

By the same argument as in the proof of Theorem 2, for each p there is a unique $v(p)$ such that $V(v(p), p) = v(p)$. Furthermore, $v(p)$ 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 $T(v(p), p) = \{0, 1\}$. This completes the proof. □

References

Bolton, P., and M. Dewatripont (2000): *Introduction to Contract Theory*. in preparation.

Caplin, A., and K. Eliaz (2002): "Aids Policy and Psychology: A Theoretical Approach," forthcoming, Rand Journal of Economics. URL: http://www.econ.nyu.edu/user/caplina/PAPERS/aids.pdf.

Caplin, A., and J. Leahy (2003): "The Supply of Information by a Concerned Expert," *Economic Journal*, forthcoming.

Christakis, N. A. (1999): *Death Foretold: Prophecy and Prognosis in Medical Care*. Chicago and London: The University of Chicago Press, second edn.

Crawford, V. P., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–1451.

Croghan, T. W. (2001): "The Controversy of Increased Spending for Antidepressants," *Health Affairs*, 20(2), 129–135.

Cutler, D. M. (2002): "Manuscript on the Medical System," In preparation; URL: http://post.economics.harvard.edu/faculty/dcutler/Book/book_draft.pdf.

Fischbach, R. L. (1983): *Physicians' Willingness to Share Information Including Medical Records Patients: Sociological, Ethical, and Legal Factors*. Ann Arbor, MI: University Microfilms International.

Grossman, S. J., and O. D. Hart (1980): "Disclosure Laws and Takeover Bids," *Journal of Finance*, 35(2), 323–334.

Hopper, S. V., and R. L. Fischbach (1989): "Patient-Physician Communication When Blindness Threatens," *Patient Education and Counseling*, 14, 69–79.

Jung, W.-O., and Y. K. Kwon (1988): "Disclosure When the Market Is Unsure of Information Endowment of Managers," *Journal of Accounting Research*, 26(1), 146–153.

Kőszegi, B. (2002): "Health Anxiety and Patient Behavior," Mimeo; URL: http://emlab.berkeley.edu/users/botond/anxiety.pdf.

——— (2003): "Health Anxiety and Patient Behavior," *Journal of Health Economics*, forthcoming.

Legorreta, A. P., J. Christian-Herman, R. D. O'Connor, M. M. Hasan, R. Evans, and K.-M. Leung (1998): "Compliance With National Asthma Management Guidelines and Specialty Care: A Health Maintenance Organization Experience," *Archives of Internal Medicine*, 158(5), 457–464.

Levine, C. (1983): "Delaying the Diagnosis: Truth-Telling and Multiple Sclerosis," *The Hastings Center Report*, 13(3), 2–3.

McBride, P., H. G. Schrott, M. B. Plane, G. Underbakke, and R. L. Brown (1998): "Primary Care Practice Adherence to National Cholesterol Education Program Guidelines for Patients With Coronary Heart Disease," *Archives of Internal Medicine*, 158(11), 1238–1244.

McGuire, T. G. (2000): "Physician Agency," in *Handbook of Health Economics*, ed. by A. J. Culyer, and J. P. Newhouse, vol. 1A, pp. 461–536. Amsterdam; New York and Oxford: Elsevier Science, North-Holland.

Milgrom, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications," *The Bell Journal of Economics*, 12(2), 380–391.

Prendergast, C., and L. Stole (1996): "Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning," *Journal of Political Economy*, 104(6), 1105–1134.

Ptacek, J. T., and T. L. Eberhardt (1996): "Breaking Bad News: A Review of the Literature," *The Journal of the American Medical Association*, 276(6), 496–502.

Quint, J. C. (1977): "Institutionalized Practices of Information Control," in *Medical Men and Their Work–A Sociological Reader*, ed. by E. Freidson, and J. Lorber. Chicago: Aldine.

Scharfstein, D. S., and J. C. Stein (1990): "Herd Behavior and Investment," *American Economic Review*, pp. 465–479.