ECONOMIC RESEARCH REPORTS

The Supply of Information by a Concerned Expert

By

Andrew Caplin and John Leahy

RR# 99-08 **April 1999**

C.V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, NY 10003-6687

Introduction

You are a doctor. You have a patient who is about to undergo medical treatment, and who knows relatively little about the procedure he is about to face and its consequences. Should you provide him with a great deal of information on the procedure, should you say nothing, or should you follow a more subtle strategy, such as providing information according to some signal of his desire for information?

This question is currently the subject of considerable interest in the burgeoning field of behavioral medicine (Morgan, Roufeil, Kaushik, and Bassett [1998]). This research suggests that the presentation of information to individuals about to undergo surgery can have a major impact both on pre-operative anxiety and on post-operative recovery. The results also indicate that the impact of information can vary radically from individual to individual, with some patients, described by the researchers as those with an avoidant coping style, appearing to suffer damage as a result of receiving “excessive” information.

> “Our findings and those of other researchers suggest that it may be clinically useful to provide information congruent with a patient’s coping style before an aversive procedure, such as colonoscopy. At the very least, endoscopists should be aware that individual patient’s may have quite different preferences in relation to precolonoscopy information.... Clinicians and researchers should also explore the ability of noncongruent information (particularly avoiders receiving sensory information) to heighten distress and result in poorer adjustment to colonoscopy because of unwanted information provision....” (Morgan et al. [1998])

These experiments and its precursors have all used questionnaire procedures to identify the personality type of the various patients. In light of the apparent value of the answers to these questionnaires, Morgan et al. [1998] make a natural policy proposal.

> “A practical solution would be to assess coping style and preferences for information before colonoscopy .... and then tailor information interventions to individual needs. .. Assessment of coping style and provision of congruent information may have economic implications for the health care system in terms of shorter postoperative recovery times and reduced need for postoperative medications and interventions....”

In this paper, we develop a simple model that allows us to explore the efficacy of this form of questionnaire-based policy. While the situation that we describe is transparent and the results are intuitive, the model itself is far from standard. We cannot rely on the classical theory of choice under uncertainty, since we need to allow for the possibility that more information may be a bad thing. We cannot rely on classical game theory, since we need to allow payoffs to depend on more information than can be represented in the game tree.

Fortunately, there are extensions to both theories that are ideally suited to our model-building task. With respect to individual decision making, we apply the “psychological expected utility” model of Caplin and Leahy [1997] (henceforth CL). With respect to the structure of the game, we apply the “psychological game” model of Geanakoplos, Pearce, and Stacchetti [1989] (henceforth GPS). In analytic terms, our biggest contribution is to open up a rich set of new interactive questions by appropriately combining these two frameworks.

The model that we develop in this paper comprises a game between a doctor and a set of patients. The patients face an impending operation. The doctor has some information about the operation that the patients do not have. The doctor is concerned with the welfare of the patients. The doctor and the patients engage in a two-stage interaction. In the first stage, the doctor asks each patient to reveal some information about their type. After this, the doctor provides information to patients as a function of their answers.

The model allows us to illustrate situations in which the simple two-stage procedure proposed by

Morgan et al. can achieve the first best solution, with information provided to those who would benefit from it, and held back from those who would not so benefit. We also illustrate several forces that may result in inefficient information provision. Where inefficiencies are identified, we investigate the possibility of using commitment devices and monetary incentives to improve upon the outcome.

The first source of inefficiency that we explore is based on the nature of the information itself. Is the information neutral, or might one piece of information be interpreted as good news, and another as bad news? We show that there is likely to be sub-optimal information supply in a "good news-bad news" scenario. The reason for this is that the doctor will not be able to resist giving the patient good news, so that no news will be bad news. A second source of inefficiency arises if the information provided by the doctor is non-verifiable. When information cannot be verified, there is a natural incentive to mislead the patient, by, for example, offering reassuring noises about a dangerous procedure. A third source of inefficiency arises if the information provided by the patient is non-verifiable. In this case, the patient may have an incentive to mislead the doctor out of a desire to influence the doctor's perception of who they are.

Our approach can be generalized to cover a wide variety of settings in which information provided by one party can directly impact another party's state of mind. Parents must make decisions concerning when to reveal "truths" about the world to their children. Teachers must decide how to express their opions and how much of their opinions to reveal to their students. Policy makers must decide what to reveal to their constituents.

In section 2 we describe the experimental literature in behavioral medicine, and set up the broad structure of our model. In section 3, we develop our model of individual preferences over lotteries. We show that the model of Kreps and Porteus [1978] is inapplicable, and that models in the spirit of CL are needed. In section 4 we show that our model lies outside classical game theory, and establish the applicability of the psychological game framework of GPS. In section 5, we solve the basic model, and show that with neutral information, the equilibrium to the game achieves the first best solution of giving an individual information if and only if it benefits them. In section 6 we allow for the information to be good or bad news, and show that this may destroy the possibility of achieving the first best. In section 7 we consider extensions to the model and to the broader analytic framework. Section 8 concludes.

Behavioral Medicine: From Experiment to Model

Medical Anxiety, Coping Styles, and Information

The interaction between psychology and medicine is a vast and rapidly growing area of study. footnote The recent experiment contributes to the analysis of stress and anxiety, which is but a small part of this broad inter-disciplinary collaboration. According to the glossary of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association [1987]), the term anxiety denotes "apprehension, tension, or uneasiness that stems from the anticipation of danger". The manifestations are said to include "motor tension, autonomic hyperactivity, apprehensive expectation, and vigilance and scanning"(p.392).

Two important elements that the above definition captures are the anticipatory nature of anxiety, and the aversive nature of the subjective experience of anxiety. The early theoretical and experimental evidence on the anticipatory and aversive nature of anxiety is summarized in Lazarus [1966]. He defines *threat* as the variable that induces anxiety, and concludes:

"Degree of threat is a function primarily of amount, imminence, and likelihood of the anticipated harm" (Lazarus [1966], p.43).

As the experimental literature has advanced, so the field has gained more information on the determinants of anxiety. footnote Breznitz [1984] analyzed the relationship between anxiety and the time remaining before the aversive event is to occur. He found a U-shaped relationship, with high anxiety when the subject is informed of the event, which diminishes until the event is very close, whereupon it increases up till time of occurrence. Monat, Averill, and Lazarus [1972] found a highly complex relationship between conditions of uncertainty and anxiety. For an aversive event that may or may not occur, the degree of anxiety depends in a non-linear way on the probability of occurrence. footnote

Psychologists have long hypothesized that patients not only suffer feelings of heightened anxiety in the face of an up-coming medical procedure, but also that this anxiety can have a significant negative effect on medical outcomes. This view has resulted in a voluminous literature on techniques for reducing the anxiety of patients prior to a stressful procedure. One of the key questions is the role of information in impacting stress. The pioneering work in this area is due to Janis [1958], who hypothesized that more information about an up-coming medical procedure would prove to be a "good thing" for patients. He believed that the information would stimulate the "work of worrying" that would initially raise anxiety, but subsequently lower anxiety and speed recovery. Following Janis, in the early days of the medical literature, the basic hypothesis was that more information would be better than less, so that the widespread dispersal of such information was seen as a policy goal. For example, Klusman [1975] reported results of an experiment in which childbirth information generally lowered anxiety, and lowered patient self reports of pain.

In following up on the Janis model, experimentalists discovered that different individuals had quite distinct responses to information. footnote The literature on heterogeneity in the desire for information has grown rapidly in the past thirty years. There is a broad consensus that it is worthwhile to distinguish at least two broad attitudes to information during the anticipatory stage of an aversive event. Roth and Cohen [1986] catalog fourteen separate literatures that use distinct but analogous definitions to capture the difference in style between those who "approach" information at this anticipatory stage and those who "avoid" information. footnote The generic term for these differences is "coping style". Research has recently been initiated on the physiological concomitants of differences in coping style. Recent evidence suggests that individuals with different coping styles may have very different physiological responses to uncertainty that are consistent with the differences in information preferences (Brown et al. [1996]).

Many researchers believe that people know their own preferences for information, and seek information only if this helps to lower their anxiety, preferring not to hear information that would raise anxiety (e.g. Miller and Mangan [1983]). Given self-awareness of information preferences, this may allow one to sort people out into those who do and those who do not want information on the basis of prior behavior. It is precisely this "revealed preference" spirit that suggested to medical researchers that it may be appropriate to use questionnaires to separate types according to information preference, and to provide information accordingly. footnote

The early experimental literature on the importance of providing information that is congruent with the patient's "coping style" is summarized in Morgan et al. [1998]. The details of the questionnaire used to separate types, the information involved, and the anxiety and welfare measures used vary from experiment to experiment. footnote Despite their heterogeneity, all of the studies make the same broad qualitative conclusions. Those whose questionnaire answers reveal a desire for information will indeed do better (in a wide variety of psychological and medical measures) if they receive more information, while those who reveal a lack of desire for information will do better if they receive less information.

"In summary, this study has shown that colonoscopy is an anxiety-provoking procedure

and that assessment of coping style and provision of congruent information significantly reduced state anxiety, recovery time, and observed behavioral indices of pain in patients undergoing colonoscopy...........
In addition to informing patients about colonoscopy, a significant goal of precolonoscopy information should be reduction in subjective anxiety." (Morgan et al. [1998]).

The Structure of the Game

We set up a two period game-theoretic model that allows us to explore the virtues and the drawbacks of the questionnaire procedure suggested by the medical researchers. The game is played between a doctor ("she") and a patient ("he") who is facing an impending operation. All moves in the game take place at the start of period 1, and the operation itself takes place in period 2. The first move is by nature, and it involves a single patient being drawn from the broader population. There are two different patient types in the overall population, and we assume that nature draws the patient from this pool at random. The precise specification of the overall patient population is the subject of section 3 below: the key point is that patient type influences the desire for information.

Once nature has selected the patient, the next move is by the patient, who chooses a message to show to the doctor. It simplifies matters without impacting essentials to assume that if the patient chooses to show something to the doctor about his type, he does so in a manner that is verifiably true. Issues associated with non-verifiable information are taken up in chapter 7 below.

Once the "show" stage is over, the game moves into its "tell" stage. Nature makes the first move in this stage, revealing information about the impending operation to the doctor, but not to the patient. For simplicity, the two possible outcomes of the operation are outcome α and outcome β: we allow the precise difference between these outcomes to vary from setting to setting. The actual outcome for an individual patient depends both on chance, and on the actual nature of the operation. In a type A operation, a fraction $\bar{p} \in (0.5, 1]$ of outcomes are of type α, while in state B the proportions are reversed, with probability $\bar{p}$ of outcome β. The common belief of all agents in an ex ante sense is that the operation has a 50% chance of being of type A, and a 50% chance of being of type B. Nature shows the doctor the true type of the operation, so that the parameter $\bar{p}$ determines the extent of the doctor's initial informational advantage over the patient.

Armed with the message from the patient and knowing the true state of the world, the doctor gets to make the final move in the game,. The doctor can tell the patient the type of the operation, A or B, or choose to reveal no information. The signal that she passes on can of course depend on what the patient revealed earlier. We also start out by assuming that the doctor's information about the operation is verifiable, consisting of some data that would be very costly to falsify, and that would expose her to considerable risk and expense if found to be false. She is empathic, and provides the information according to the impact on patient utility.

It remains only to provide details of the utility functions of the various patient types, and of the equilibrium concept for the game. Neither is standard.

Modeling Individual Preferences

Why not Kreps-Porteus Preferences?

We assume that all patients have a classical expected utility function over second period outcomes, with outcome α at least as good as outcome β, $U_2(\alpha) \geq U_2(\beta)$. We normalize by setting $U_2(\beta) = 0$. We let p_2 denote the probability of outcome α as opposed to β as seen from the start of period 2. If we were to assume classical expected utility preferences, the first period ranking of all lotteries over period 2 outcomes would be trivially determined. However, we wish to allow the manner in which uncertainty unfolds to influence patient preferences.

It is natural to hypothesize that the appropriate framework is the model of preference for early as against late resolution of uncertainty of Kreps and Porteus [1978a]. To formalize their approach to our medical problem, we let $Z_2 = \{\alpha, \beta\}$ denote the second period pure prize space, and define D_2 to be the set of second period probabilities over Z_2, with generic element $d_2 \in D_2$. We then let D_1 denote the space of probability measures over these lotteries, with generic element $d_1 \in D_1$. These are referred to as “temporal lotteries”. Rather than assume a full substitution axiom applies to D_1, they assume a less comprehensive “within period” variant of the substitution axiom that yields their basic representation theorem. We refer to the axioms as the KP axioms.

Theorem (*Kreps and Porteus* 1979, *Proposition* 1) *The KP axioms are necessary and sufficient for there to exist two continuous functions, $U_2 : Z_2 \to R$ and $u_1 : R \to R$, with u_1 strictly increasing, such that if $U_1 : D_2 \to R$ is defined by,*

$$U_1(d_2) = u_1(E[U_2(z_2)|d_2]),$$

then $d_1 \succ d_1'$ if and only if $E[U_1(z_1)|d_1] > E[U_1(z_1)|d_1]$. footnote

The monotonicity of the u_1 function implies that there is no reversal of preferences over period 2 lotteries, so that preferences over lotteries are time consistent. However if u_1 is not linear, a first period lottery over lotteries over second period prizes cannot be reduced to the simple compound lottery over second period prizes. This gives the theory room to allow preference either for the early or for the late resolution of uncertainty. The KP axioms allow some people to prefer knowing how happy they will be tomorrow, while others prefer not to know how happy they will be tomorrow. footnote

At first sight, it appears that allowing for patients with Kreps-Porteus preferences will successfully formalize the doctor’s choice problem. The doctor who knows the true state of the world first finds out whether the patient is resolution loving or resolution averse. If the patient is resolution loving, she provides the information, since this patient prefers ignorance to knowledge. If the patient is resolution averse, she provides no information, since the patient prefers knowledge to ignorance. Note that if this strategy is known to the patient, he has every reason to reveal his type to the doctor, since the end result is ignorance for the patient who prefers ignorance, and knowledge for the patient who prefers knowledge.

It turns out that the above logic is incomplete. The issue that it sweeps under the rug is the need to induce the doctor to reveal the information once it is in her hands. The subtle point here is that the doctor must be able to decide what to do in the face of a superior understanding of the lottery that the patient is about to face. The decision on whether or not to provide information at this stage must be based on an assessment of whether or not to leave the patient with an illusion about tomorrow’s lottery. The Kreps-Porteus model is not rich enough to cover preferences over illusions.

Consider a special case of our model with $\bar{p} = 1$. Suppose that the doctor knows that the operation is of type A and that she knows the two relevant functions $U_2 : \{\alpha, \beta\} \to R$ and $u_1 : R \to R$ that define the patient’s preferences. On what basis can she decide whether or not her patient would be better off if being informed of the true state rather than being left ignorant?

The situation is transparent if she reveals the information to the patient. In this case, the patient immediately knows that their prize tomorrow is α for sure, and the doctor knows that the patient’s utility is $u_1(U_2(\alpha))$. We must compare this with the doctor’s belief about the overall utility of not revealing the information to the individual in state A. The problem is that the doctor’s beliefs about the patient’s utility in this case are ill-defined.

One way for the doctor to view the situation is to resolutely maintain the patient’s viewpoint. In period 1, the uninformed patient views outcome α as equiprobable to outcome β, and therefore assigns the lottery the utility $u_1\left(\frac{U_2(\alpha)+U_2(\beta)}{2}\right)$. footnote The problem with sticking with the patient’s false view of the situation is that the doctor really does know better, and she knows that the patient’s state of mind is based on an illusion. She has no reason to believe that it is in the patient's best interest for her to ignore her information. footnote

This suggests that the appropriate way for the doctor to proceed is to use her superior knowledge in assessing the lottery. The doctor knows for sure that the patient is going to receive prize α in period 2. If she takes this view of the lottery that the patient is facing, then she is forced to conclude that, in her view, the patient has the same utility regardless of whether or not he is informed. The fallacy involved in the doctor taking this view is that it incorrectly attributes to the patient her own certainty about what the ultimate prize is going to be.

In technical terms, the limitation of the model is that it does not allow one to separate out the first and the second period viewpoints in a manner that captures the doctor's awareness that the patient, if uninformed, is living in illusion. To capture this, the doctor must assess patient utility in a case in which the patient believes that α and β are equiprobable in period 1, and yet the actual outcome in period 2 is α with probability 1. The space of temporal lotteries is simply insufficiently rich to allow us to ask, let alone answer, questions involving preferences over illusions.

Psychological Expected Utility Theory and Anxiety

Our resolution to this apparent difficulty is to return to an expected utility framework and to enrich the prize space to include the anticipatory phase, as proposed in CL. Provided we are absolutely explicit about all prizes in the lottery, the doctor will have no difficulty in working out her beliefs about patient utility if she chooses to leave him uninformed.

With respect to period 2, there are still just the two pure prizes, $Z_2 = \{\alpha, \beta\}$. Following CL, we add an additional anticipatory prize based on the mental state in the first period. We refer to the first period prize as the level of anxiety, $X \geq 0$. We assume that a standard substitution axiom applies to prize vectors comprising first period anxiety and the actual outcome of the operation in period 2. We assume that this expected utility function is separable between periods. We measure units of anxiety in such a way that the first period expected utility function is linear, with higher levels of anxiety being more aversive,

$$U_1(X) = -X.$$

All patients have the same utility function over anticipatory and actual prizes, but they may have different internal processes for producing anxiety based on beliefs about the future. We assume that the actual level of anxiety in period 1 depends on p_1, the probability in the patient's mind that the second period prize will be α rather than β. Each particular first period belief gives rise to a deterministic level of anxiety, $X(p_1)$. Once we have modeled the connection between the first period belief and the corresponding level of anxiety, the expected utility nature of preferences allows us to extend preferences to the entire space D_1 of temporal lotteries.

PEU and Preferences over the Resolution of Uncertainty

We allow for heterogeneity in the determinants of anxiety. Specifically, we allow preference for early resolution of uncertainty as well as preference for late resolution of uncertainty in the spirit of Kreps and Porteus. We assume that the patient population is divided into two types, $\tau = E, L$. In the population as a whole, 50% of individuals are of type E. For patients of type E (early resolvers) feelings of anxiety are more severe the less certain they are about the outcome in period 2, while individuals of type L (late resolvers) experience higher levels of anxiety the more certain they are about the outcome.

The distinction between E and L types is determined by the shape of the anxiety function, $X(p_1)$. Given two distinct temporal lotteries d_1, d_1', the patient of type E experiences less anxiety if he knows today which lottery he will face rather than having to wait until tomorrow to find out. This will follow if and only if the anxiety function $X(p_1)$ is concave in p_1. Conversely, a patient of type L always prefers to be uncertain about which lottery he will receive. In our framework, this follows if and only if the production function for anxiety is convex in p_1.

We provide a simple symmetric functional form for the anxiety function. For a type L patient, the production function for anxiety takes the form,

$$X^L(p_1) = (p_1 - \frac{1}{2})^2.$$

For a type E patient, the production function takes the form,

$$X^E(p_1) = -(p_1 - \frac{1}{2})^2.$$

These functions have the appropriate shape. They are also mirror images of one another.

Putting the second period expected utility function over pure prizes, $U_2(\alpha) \geq U_2(\beta) = 0$, together with the production function and the utility function for anxiety allows us to derive the overall expected utility functions over period 1 beliefs for type $\tau = E, L$. Following CL [1997] we refer to this as the induced expected utility function over period 1 beliefs, which we denote $V^\tau(p_1), \tau = E, L$. The induced expected utility function for an E type is,

$$V^E(p_1) = (p_1 - \frac{1}{2})^2 + p_1 U_2(\alpha)$$

The induced expected utility function for an L type is,

$$V^L(p_1) = -(p_1 - \frac{1}{2})^2 + p_1 U_2(\alpha).$$

Preferences get extended to the space of temporal lotteries using the expected utility property. Given $d_1, d_1' \in D_1$, the agent of type $\tau = E, L$ has $d_1 \succeq d_1'$ if and only if,

$$E[V^\tau(p_1)|d_1] \geq E[V^\tau(p_1)|d_1'].$$

Note that the overall function $V^\tau(p_1)$ comprises two quite distinct components, which correspond to the two different periods. The linear portion of the function relates to the second period expected utility function over outcomes. The quadratic portion relates to the first period expected utility function over levels of anxiety. The fact that the preference over lotteries comprises these qualitatively different period-by-period components gives rise to a characteristic feature of CL [1997], which is time inconsistency of preferences. In a richer game, this time inconsistency would play a critical role. In the current setting it is unimportant, since all decisions are made at the start of period 1.

Given this expected utility model of preferences, the doctor is perfectly able to calculate patient utility in cases in which the anxiety prize is determined by a first period belief p_1 that the doctor knows to be false. Return to the case in which the doctor knows that the operation is sure to have outcome α, and knows also that the patient starts off with the belief that it is equally likely to be α or β. If the patient learns nothing in the first period, the doctor assesses patient anxiety as $X^\tau(\frac{1}{2})$, while the second period prize is α with probability 1. Substitution in the patient's expected utility function allows the her to assess patient expected utility as,

$$U_1(X^\tau(\frac{1}{2})) + U_2(\alpha) = -X^\tau(\frac{1}{2}) + U_2(\alpha).$$

The simplicity of this calculation is quite unaffected by the doctor's knowledge that the patient is living with an illusion in period 1. The fact that the patient's beliefs in period 1 are misplaced does not prevent them from being crucial to welfare in that period. However, in period 2 it is only the true prize that influences the payoff, so that the doctor uses her superior knowledge of this lottery to assess patient welfare. It is this ability to model preferences over illusions that accounts for the applicability of the expected utility approach of CL [1997]. footnote

The Doctor-Patient Game as a Psychological Game

Why Beliefs about Strategies Influence Payoffs

Figure 1 presents the extensive form of this game. The tree begins with the move by nature (*N*) in which patient type (*E* or *L*) is determined. The patient (player 1) then chooses whether or show (*S*) or not show (*NS*) his type, whereupon the doctor (player 2) learns the true nature of the operation. The doctor makes the final choice of whether or tell (*T*) or not tell (*NT*) the patient the true state. The curved lines connect nodes that belong to the same information set.

The payoffs to strategies involving revelation by both doctor and patient are easily determined. If a patient shows that he is of type *E*, then the payoff to both doctor and patient from doctor revelation is $V^E(\bar{p})$. If a patient reveals that he is of type *L*, then the payoff to both agents from doctor revelation is $V^L(\bar{p})$. Patient payoffs can also be trivially computed whenever the doctor reveals information, even when the patient follows the strategy of non-revelation.

The complexity of the game lies in the doctor's payoff in cases in which the patient has chosen to not reveal his type, and in both the doctor's and the patients' payoffs in the cases in which the doctor does not reveal the true state. These payoffs cannot be specified without additional information on the strategies of the two players. In figure 1 we use asterisks to represent the various strategy-dependent payoffs. Where there is one asterisk, the payoff to the agent in question depends directly on their belief about the strategy of the other. Where there are two asterisks, the payoff to the agent depends on their belief concerning the belief of the **other** agent about **their** strategy.

To understand what determines the payoffs in all asterisked positions, consider first the doctor's payoff if the patient does not reveal his type, all of which have a single asterisk. The payoff to the doctor of revealing information to a patient who has not revealed his type depends on her belief about the patient's type. To assess this, she uses her beliefs about the strategy that the patient pursued at the earlier stage of the game. Similarly, the payoff to the patient when the doctor does not reveal any information about the state depends on his belief about what doctor non-revelation implies for the true state of the world, which is determined by the doctor's strategy.

The cases with double asterisks all represent payoffs to the doctor in cases in which she does not reveal information to the patient. The doctor is aware that the patient's beliefs will depend on his interpretation of her strategy of non-revelation. Being empathic, she will need to form her own beliefs about what the patient believes in order to compute her estimate of patient expected utility, which is also her own utility.

Psychological Games and Psychological Equilibria

The dependence of the final payoff on beliefs about strategies moves this model away from classical game theory. In this subsection we introduce the general two player extensive form psychological games of GPS, mapping our game into their apparatus in the next subsection.

The GPS model begins with a classical game tree model with perfect recall. The set of terminal nodes of the tree is denoted *T*, with generic element $t \in T$. As usual there is common knowledge about nature's moves, which are determined by some fixed system of probability distributions, ρ.

Other standard elements of the model are the set of player i's behavior strategies, Σ_i, and the overall space of behavioral strategies, $\Sigma = \Sigma_1 \times \Sigma_2$, with generic element $\sigma \in \Sigma$.

What distinguishes this as a psychological game as opposed to a classical game is that the two players' utility functions depend not only on which of the possible pure strategy profiles are followed, but also on relevant beliefs. GPS refer to the overall domain of beliefs as the set of collectively coherent initial beliefs, $\bar{B} = \bar{B}_1 \times \bar{B}_2$. footnote The set $\bar{B}_i$, the coherent belief hierarchy of player i, summarize beliefs of all orders about the other player, including beliefs about the other player's strategy, their beliefs about the other player's beliefs about their own strategy, and so on.

The utility function of player i has the general form $u_i : \bar{B}_i \times T \to R$. Note that each strategy profile $\sigma \in \Sigma$, together with nature's probability distributions, ρ, induce a probability distribution over terminal notes. The utility function u_i is extended to $u_i : \bar{B}_i \times \Sigma \to R$ using the expected utility property. Taken together, the extensive form game tree, the beliefs about nature's moves, and the utility functions determine a particular extensive psychological game, Γ.

Given any coherent belief system $\bar{b} = (\bar{b}_1, \bar{b}_2) \in \bar{B}$, we reserve the notation $\Gamma(\bar{b})$ to denote the standard extensive form game with final payoffs determined as $u_i(\bar{b}_i, t)$. A very special class of beliefs are those in which it is common knowledge that a particular strategy profile $\sigma \in \Sigma$ is being played. We denote this special class of beliefs as $\beta(\sigma) \in \bar{B}$. In order for a particular strategy profile to be a psychological Nash equilibrium, it must be optimal to pursue that strategy in a setting in which the payoffs are determined using common knowledge that the strategy is being pursued.

Definition *A pair* $(\hat{b}, \hat{\sigma}) \in \bar{B} \times \Sigma$ *is a psychological Nash equilibrium of the psychological game G if and only if* (1) $\hat{b} = \beta(\hat{\sigma})$; *and* (2) *for each player i, and for each* $\sigma_i \in \Sigma_i$,

$$u_i(\beta(\hat{\sigma}), \hat{\sigma}_i, \hat{\sigma}_{-i}) \geq u_i(\beta(\hat{\sigma}), \sigma_i, \hat{\sigma}_{-i})$$

Finally, note that the standard objections to certain Nash equilibria will present themselves in this setting just as in standard game theory. Specifically, we will need to rule out "incredible" threats. Following Kreps and Wilson, a belief system $\mu \in M$ associates with each information set a probability distribution over the nodes of that set.

Definition *The triple* $(\hat{b}, \mu, \hat{\sigma})$ *constitute a sequential psychological equilibrium of the psychological game* Γ *if and only* $(\hat{b}, \hat{\sigma})$ *is a psychological Nash equilibrium of* Γ, *and* $(\mu, \hat{\sigma})$ *is a sequential equilibrium of* $\Gamma(\beta(\hat{\sigma}))$ *in the standard sense*.

The Doctor-Patient Game is a Psychological Game

In order to apply this apparatus to the model at hand, we must first show that our game can be represented as an extensive form psychological game. Since the extensive form game tree has already been presented in figure 1, and nature's moves are assumed to be common knowledge, the key is to specify the two utility functions, $u_i : \bar{B}_i \times T \to R$, $i = 1, 2$. Before doing this it is helpful to introduce some more specific notation for the strategies of the two players.

An overall player 1 (patient) strategy is a vector $r = (r^E, r^L) \in [0,1] \times [0,1]$ that specifies for each type the probability that he will show the doctor his preference type. We can summarize the strategy of player 2 (the doctor), π, by the corresponding mixed strategy profile of information revelation, depending on the true nature of the operation, $\theta = A, B$, and whether the patient is among those who are revealed to be of type E, among those revealed to be of type L, or among those who chose not to show their type: we refer to this group as type $\tau = NS$. We let π_θ^τ denote the probability that the strategy accords to revealing that the true state is θ to individuals who are of type τ, and define π to be a complete strategy,

$$\pi = (\pi_A^E, \pi_B^E, \pi_A^L, \pi_B^L, \pi_A^{NS}, \pi_B^{NS}) \in [0,1]^6.$$

We now complete the specification of patient utility in figure 1, beginning with those places in which patient utility is marked with an asterisk. These are the situations in which the doctor chooses not to reveal the true state of the operation. In this case the patient's utility depends only on the relative likelihood that he assigns to non-revelation signifying that the true state is A, which depend on his beliefs about the doctor's strategy. Looking at any hierarchy of patient beliefs, $\bar{b}_1 \in \bar{B}$, we extract from them the first order beliefs about the strategy that the doctor is pursuing, which we denote $\tilde{\pi}(\bar{b}_1)$. Looking at this strategy, it is trivial to compute the relative likelihood of a patient of type τ believing that doctor non-revelation implies that the true state is A as $\frac{\tilde{\pi}_A^\tau(\bar{b}_1)}{\tilde{\pi}_A^\tau(\bar{b}_1)+\tilde{\pi}_B^\tau(\bar{b}_1)}$, and the payoff to the patient can be trivially computed as a result. The only situation that this formula leaves indeterminate is $\tilde{\pi}_A^\tau(\bar{b}_1) = \tilde{\pi}_B^\tau(\bar{b}_1) = 1$ for some type $\tau = E, L, NS$, in which case the patient has no initial theory as to why the doctor said nothing. We discuss this case below.

We now specify the utility function of the doctor in the four branches in figure 1 in which it is marked with a single asterisk. These all involve non-revelation by the patient followed by revelation by the doctor, in which case the doctor's utility depends only on her belief about the probability that a non-revealing patient is of type E rather than of type L. Looking at the doctor's hierarchy of beliefs, $\bar{b}_2 \in \bar{B}$, we extract her first order beliefs about the strategy that the patient is pursuing, which we denote $\tilde{r}(\bar{b}_2)$. Looking at this strategy, it is trivial to compute the probability that a non-revealing patient is of type E rather than of type L as $\frac{1-\tilde{r}^E(\bar{b}_2)}{2-\tilde{r}^E(\bar{b}_2)-\tilde{r}^L(\bar{b}_2)}$, and the payoff to the doctor can be trivially computed as a result. The only situation that this formula leaves indeterminate is if $\tilde{r}^E(\bar{b}_2) = \tilde{r}^L(\bar{b}_2) = 1$, in which case the doctor has no initial theory as to how patient non-revelation occurred. We discuss this case below.

Another set of payoffs that need to be specified are those in the eight branches in figure 1 in which the doctor's payoff is marked with a double asterisk. These all involve non-revelation by the doctor. Here her payoff depends on her belief about what the patient will believe if he is not told the true state. Given a belief hierarchy $\bar{b}_2 \in \bar{B}$, we extract the doctor's second order beliefs that specify what she believes the patient believes to be her strategy, and use these to compute her estimate of patient payoff, which is also her payoff. This works in all situations except those in which she believes that the patient was certain that she would reveal the true state to him, which leaves him baffled when she does not.

The unspecified payoffs at this point involve situations that are initially regarded as inconsistent with prior beliefs. What does the patient believe to be the true state if no state is announced when he was sure that the doctor was going to announce the state? We will not need to pinpoint the exact beliefs associated with this outcome, but we will insist that it is common knowledge how this will be interpreted. We let $\lambda_A \in [0,1]$ denote the commonly understood probability that the patient assigns to state A in cases in which he was certain that the true state would be revealed. In a symmetric manner, we let $\lambda_E \in [0,1]$ denote the commonly understood probability that the doctor assigns to the patient's being of type E in cases in which she was certain that patient would reveal his type, but he did not. With these new data, we are able to specify all remaining payoffs, and complete the specification of our psychological game.

Note that we treat λ_A and λ_E as part of the commonly understood data of the problem, in just the same spirit as we treat the beliefs about nature's moves as common knowledge. Given that these are treated as data by the players, we are connecting the analysis of out-of-equilibrium beliefs that are part of the sequential equilibrium calculation with the payoff in the psychological game. Given that the model is unable to provide any particular grounds for a specific value of these parameters, it is fortunate that the solution to the model is essentially unaffected by these parameters. footnote

Solving the Basic Model: Neutral News and

Efficiency

The Model

We solve for the psychological sequential equilibria for the special case of our model in which the two pure prizes are indifferent,

$$U_2(\alpha) = U_2(\beta) = 0$$

We call this the case of **neutral news**, as distinct from the case of valenced news that we analyze later. This allows us to simplify the induced expected utility functions of the agents. The induced expected utility function for an E type is,

$$V^E(p_1) = (p_1 - \frac{1}{2})^2.$$

The induced expected utility function for an L type is,

$$V^L(p_1) = -(p_1 - \frac{1}{2})^2.$$

Note that these functions have especially strong symmetry properties due to the indifference between the pure prizes.

The Solution

We solve for the set of sequential psychological equilibria of our game, which we denote $S(\Gamma)$. In order to do this, we can limit attention to a very special class of the games $\Gamma(r,\pi)$ in which the beliefs of both players of all orders reduce to the strategy profile common strategy profile (r,π). The proof that some strategy profile $(\hat{r},\hat{\pi})$ is a sequential psychological equilibrium to the game involves simply looking at the corresponding extensive form game $\Gamma(\hat{r},\hat{\pi})$ and proving that $(\hat{r},\hat{\pi})$ is a sequential equilibrium of this game, in the standard sense.

To gain insight into the solution, figure 2 summarizes the extensive form of the game $\Gamma(r,\pi)$ with $r = (\frac{1}{3}, \frac{2}{3})$ and $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3})$. In this game the E type reveals with probability $\frac{1}{3}$, the L type with probability $\frac{2}{3}$, and the doctor reveals the true state with probability $\frac{1}{3}$ to those who announce E, with probability $\frac{2}{3}$ to those who announce L, and with probability $\frac{1}{2}$ to those who choose not to reveal their type, NS. To simplify the figure we define $\tilde{p}$,

$$\tilde{p} = \frac{2}{3}\bar{p} + \frac{1}{3}(1-\bar{p}) = \frac{1}{3}(1+\bar{p}).$$

To understand the payoffs in the figure, consider first the doctor's payoff if the patient does not reveal type. Given the patient strategy $r = (\frac{1}{3}, \frac{2}{3})$, the doctor believes that there are twice as many non-revealers of type E than of type L. The payoff to the doctor can be computed using the appropriate weighted average of the utility functions of these two types.

Now consider the payoffs to the patient when the doctor does not reveal any information about the state. For a patient who reveals his type, the doctor's non-revelation is entirely uninformative about the true state, since revelation is equally likely regardless of whether the true state is state A or state B. But for a patient who chooses not to reveal his type, the doctor is twice as likely to reveal nothing if the true state is A, and the patient appropriately assumes that non-revelation corresponds to true state A with probability $\frac{2}{3}$, with the corresponding payoff. With this we can substitute the appropriate beliefs to determine all patient payoffs in the face of non-revelation by the doctor.

The final set of payoffs to explain are those that were marked with the double asterisk in figure

1, which represent doctor payoffs in cases in which the doctor does not reveal information to the patient. If the doctor does not tell a patient who is of revealed type *E* the true state, then she knows that the patient will believe that both states are equally likely, and she will derive empathic utility $V^E(\frac{1}{2})$. Similarly if she does not tell a patient of revealed type *L* the state, then she knows that the patient will derive utility $V^L(\frac{1}{2})$.

In this example, the most intricate situation involves the doctor not telling the state to a patient who has not revealed his type. In this case, the patient revelation strategy results in a $\frac{2}{3}$ probability that the agent is of type *E*, while the doctor non-revelation strategy leaves the patient believing that there is a $\frac{2}{3}$ probability that the true state is *A*. The doctor thus derives empathic expected utility of amount Ψ,

$$\Psi \equiv \frac{2}{3}V^E\Big(\frac{2}{3}\bar{p} + \frac{1}{3}(1-\bar{p})\Big) + \frac{1}{3}V^L\Big(\frac{2}{3}\bar{p} + \frac{1}{3}(1-\bar{p})\Big).$$

To check whether the strategies $r = (\frac{1}{3}, \frac{2}{3})$ and $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3})$ constitute a sequential psychological equilibrium, we look at the game in figure 2 as a standard game of incomplete information, and solve for the sequential equilibria of the game. Note that when the patient chooses to reveal, we enter two proper subgames of the overall game. The Nash equilibria of these games are trivial. It is a dominant strategy for the doctor to reveal to the *E* type, since her payoff from revelation is $V^E(\bar{p}) = V^E(1-\bar{p}) > V^E(\frac{1}{2})$. It is a dominant strategy for her not to reveal to the *L* type, since her payoff from revelation is $V^L(\bar{p}) = V^L(1-\bar{p}) < V^L(\frac{1}{2})$. With respect to the group *NS*, the doctor's dominant strategy is to tell the patient the truth, since $\frac{2}{3}V^E(\bar{p}) + \frac{1}{3}V^L(\bar{p}) > \Psi$.

The backward induction is immediate, with the *E* types being indifferent between revelation and non-revelation of type, since both type *E* and type *NS* get the higher payoff associated with doctor revelation. The *L* types get strictly higher payoff from not being told the state by the doctor, and hence choose to reveal type. In the end, we can compute the set of sequential psychological equilibria of this game as,

$$S(r,\pi) = \{(r^*,\pi^*) = \{(q,1),(1,1,0,0,1,1)\} | q \in [0,1]\,\}.$$

Since the strategy $r = (\frac{1}{3}, \frac{2}{3})$ and $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3})$ is not in this set, we conclude that this is not a sequential psychological equilibrium. footnote

The above analysis points out how one can solve for the equilibria of the game. Since patients of type *E* are better off with the information revealed at once, sequential rationality implies that the doctor always sets $\hat{\pi}^E = (\hat{\pi}^E_A, \hat{\pi}^E_B) = (1,1)$. Since the patients of type *L* are better off not knowing, sequential rationality implies that the doctor always sets $\hat{\pi}^L = (\hat{\pi}^L_A, \hat{\pi}^L_B) = (0,0)$. Given this, it is immediate that there is an equilibrium with $\hat{r} = (1,1)$, $\hat{\pi}^E = (1,1)$, and $\hat{\pi}^L = (0,0)$.

Note that this equilibrium calls for some analysis of out-of-equilibrium beliefs. The doctor must make a decision on whether or not to reveal information to the *NS* types, even though in equilibrium this group is empty. Her decision is based on the parameter $\lambda_E \in [0,1]$ that specifies common beliefs about patient type in this situation. Her optimal strategy is to tell this group the true state if $\lambda_E > 0.5$, to say nothing if $\lambda_E < 0.5$, and to mix arbitrarily if $\lambda_E = 0.5$. The decision has no impact on the equilibrium. In fact the entire solution to the model is independent of the value of $\lambda_A,\lambda_E \in [0,1]$.

A moment's thought shows that there are many equilibria in the psychological game all of which are equivalent in terms of patient information revelation, with all *E* types learning the state, and all *L* types being told nothing. In all of these other equilibria, the *NS* group is either all of type *E* or all of type *L*. The doctor tells this group the state if they are of type *E*, not if they are type *L*, and a patient of this type mixes in any way between revealing and not revealing his type.

Proposition *The set of sequential psychological equilibria of* Γ *is,*

$$S(\Gamma) = \{(q,1),(1,1,0,0,1,1)|q \in [0,1]\} \cup \{(1,q),(1,1,0,0,0,0)|q \in [0,1]\}.$$

In all of these equilibria, all L type patients get payoff $V^L(\frac{1}{2})$*, while E type patients get payoff* $V^E(\bar{p})$ *if the true state is A, and* $V^E(1-\bar{p})$ *if the true state is B.*

Note that the doctor is happy with the equilibria in this game, since they all achieve her first best expected utility,

$$\frac{1}{4}V^E(\bar{p}) + \frac{1}{4}V^E(1-\bar{p}) + \frac{1}{2}V^L(\frac{1}{2}).$$

It is also worth noting that there is a more passive game form that would achieve the same outcome. The doctor could simply set up a room, and make it known that information on the true state would be provided only to those who entered the room. The type E patients would chose to enter the room, while the type L patients would not. Both the efficiency of the equilibrium and the equivalence to a game in which the doctor is passive are very special features of the model with non-valenced information. In fact, the passive strategy may not be feasible in certain cases.

Valenced Information and Inefficiency

Valenced Information and Anticipatory Utility

Doctors typically have far deeper knowledge of many aspects of the operation than do patients, including the difficulty of the operation and the prognosis for a complete recovery. We would not expect patients to be indifferent about these outcomes. We therefore move to the case of valenced information, $U_2(\alpha) > U_2(\beta) = 0$, and normalize to $U_2(\alpha) = 1$.

The psychological literature on anxiety provides compelling evidence that patient beliefs about the unpleasantness of an up-coming procedure impacts patient welfare. To capture this, we assume that in addition to the negative effect of anxiety, there is a positive effect on patient welfare of believing that the outcome tomorrow is more likely to be the better outcome. For analytic simplicity, we assume that this effect is separable from the anxiety prize in the utility function, and refer to it as the anticipated utility prize.

We measure the patient's anticipated utility prize by $F \geq 0$, with higher levels corresponding to a more optimistic period 1 vision of the lottery in period 2. The representative agent's overall period 1 utility function is,

$$U_1(X,F) = -X + F.$$

We assume that the anticipated utility is produced in a trivial manner, in that it depends in a simple linear and increasing fashion on the expected utility from the operation in period 2 as assessed in period 1,

$$F(p_1) = \gamma EU_2(p_1) = \gamma p_1.$$

Here $\gamma > 0$ is a scaling factor that weights the impact on preferences of the foreshadowing relative to the anxiety. footnote

As before the population is divided into the two types, $\tau = E, L$, with equal numbers of each type E. For type E individuals, anxiety is produced as $X^E(p_1) = -(p_1 - \frac{1}{2})^2$. For type L patients, anxiety is produced as $X^L(p_1) = (p_1 - \frac{1}{2})^2$. Combining the simple linear expected utility function with the linear production function for F, and the quadratic production function for X allows us to trivially derive the period 1 induced expected utility function over beliefs for type $\tau = E, L$ as,

$$V^E(p_1) = \gamma p_1 + (p_1 - \frac{1}{2})^2 + p_1 U_2(\alpha),$$
$$V^L(p_1) = \gamma p_1 - (p_1 - \frac{1}{2})^2 + p_1 U_2(\alpha).$$

No News is Bad News

The form of the game tree for this new psychological game is unchanged. The only difference is in the payoff structure. Consider the upper-most branch of the game tree in figure 1, in which a type E patient shows the doctor his type, whereupon she shows him that the true state is A. In this case both patient and doctor receive the prize $V_1^E(\bar{p}) + \bar{p}$. All other prizes involving show by the patient and tell by the doctor are equally easy to compute. Prizes in all remaining branches are strategy and belief-dependent, but are easy to compute in the precise spirit of the original game of the last section. We refer to the new psychological game that results as Γ^{VA}.

One conjecture to consider is that the game may have precisely the same equilibria as the original game, with information about the operation revealed to those who show that they are of type E, and held back from those of type L. Consider the typical equilibrium of the original game, with both patient types showing the doctor who they are, $r = (1,1)$, and the doctor revealing the state to the type E, not to the type L, $\pi^E = (1,1), \pi^L = (0,0)$. For simplicity, we assume that $\lambda_E > 0.5$, so that the doctor chooses to tell a patient of type NS the truth in this equilibrium, $\pi^{NS} = (1,1)$, on the implicit assumption that he more likely to be of type E than of type L. This is not important to the analysis.

To check whether or not these strategies constitute an equilibrium to the new game, we begin by identifying the sequentially rational strategies. Given her concern with patient welfare, the doctor's informational strategy must be to provide information optimally (given her view of the operation) to those who have revealed their type, and to reveal information or not to the non-announcers according to her rational beliefs about the constitution of this group.

The key situation to check involves the combination of a patient who has revealed that he is of type L, and a doctor who knows that the operation is of type A. If the doctor does not tell the patient the state, the doctor gets payoff $V_1^L(\frac{1}{2}) + \bar{p}$. If the doctor does tell the patient, she gets payoff $V_1^L(\bar{p}) + \bar{p}$. It is rational for her not to reveal the information if and only if,

$$V_1^L(\bar{p}) = \gamma\bar{p} - (\bar{p} - \frac{1}{2})^2 \leq \frac{\gamma}{2} = V_1^L(\frac{1}{2}) \Leftrightarrow \gamma \leq (\bar{p} - \frac{1}{2}).$$

In fact, it is clear that the equilibrium set for Γ^{VA} is the same as that for Γ if this inequality is strict.

In cases with $\gamma > (\bar{p} - \frac{1}{2})$ the doctor knows that even the type L patient's utility will be higher knowing that the news is good than remaining ignorant. Of course, the patient's limited information produces the belief that he would prefer to remain ignorant. The subtle point here is that while the patient of type L believes that he **does not** want the information, the doctor knows that he **does**. It is this difference of opinion that accounts for the doctor's decision to reveal the good news to the L type.

Given the symmetry of the model, this same condition on the parameters, $\gamma > (\bar{p} - \frac{1}{2})$, ensures that if the operation is of type B, then the doctor knows that the E type does not want to know. One might naively think that this would imply that the doctors must choose to not reveal information to the type E patients if the true state of the world is B. Of course it is not quite this simple. The non-revelation of information is only partly a matter of choice, since the interpretation of being told nothing is endogenous. In this particular case, since the patient is well aware that he would have been told something had the operation been of type A, it is trivial for him to conclude that silence implies that the true state of the world is of type B.

Overall it is natural to conjecture that in cases with $\gamma > (\bar{p} - \frac{1}{2})$, all equilibria to the game will involve both types of patient knowing the true state for sure. In fact, this is simple to confirm. With this condition on the data, we have ensured via sequential rationality that the doctor will reveal that the true state is *A* not only to type *E* patients, but also to type *L* and *NS* patients. At this point, the rest of the strategies are irrelevant. Since the patient can be sure that they will be told if the true state is *A*, the two alternative strategies for the doctor of either remaining silent or of admitting that the state is *B* are informationally equivalent. The patient will know that unless they are told that the world is in state *A*, the world is in state *B* for sure. Proceeding with the backward induction, the patient will be indifferent concerning whether or not to show the doctor his type, since the information revealed to all patients is identical.

Proposition *With $\gamma > (\bar{p} - \frac{1}{2})$, the set of sequential psychological equilibria of Γ^{VA} is,*

$$S(\Gamma^{VA}) = \{(r^E, r^L), (1, \pi_B^E, 1, \pi_B^L, 1, \pi_B^{NS}) | r^E, r^L, \pi_B^E, \pi_B^L, \pi_B^{NS} \in [0,1]\}.$$

In all of these equilibria, all patients of type $\tau = E, L$ get payoff $V^\tau(\bar{p})$ if the true state is A, and $V^\tau(1-\bar{p})$ if the true state is B.

Neither the doctor nor the *L* type patients are happy with this equilibrium. The doctor's overall expected utility is,

$$\frac{1}{4}V^E(\bar{p}) + \frac{1}{4}V^E(1-\bar{p}) + \frac{1}{4}V^L(\bar{p}) + \frac{1}{4}V^L(1-\bar{p}) < \frac{1}{4}V^E(\bar{p}) + \frac{1}{4}V^E(1-\bar{p}) + \frac{1}{2}V^L(\frac{1}{2}).$$

It is clear that both the doctor and the type *L* patient would be better off if the doctor could commit to not revealing the truth in state *A*. Unfortunately, the doctor is not able to commit to not providing information to the *L* types in this game form, and this inability to commit to the ex ante optimal strategy produces the welfare problem.

Changing the Game

One attempt to provide a commitment mechanism would be for the doctor to follow the passive strategy alluded to in section 5.3 above, with patient self selection deciding the issue of whether or not the patient learns the true state. This achieves the first best, since the *L* type bases his decision not to become informed on his ignorant prior. But can the doctor really withhold information in this way? Once the patient has chosen not to find out the truth, can the doctor be constrained from handing on any good news that he thereby missed? If so, why is she withholding, since she is needlessly damaging the patient? If not, how can she prevent no news from being interpreted as bad news?

When and how it may be feasible to commit to not provide information is an open question. In our experience the commitment problem is real, and doctors use a wide variety of (largely ineffective) techniques for attempting to leave patients who have bad news in an optimistic frame of mind for as long as possible.

The prototypical case involves a patient who has a check-up for a possible medical problem. The test arrives at the doctor's office, but the doctor is unavailable for one week. What will happen if the patient phones the nurse? Typically, if the news is good, the nurse lets slip some phrase such as "not to worry. I can see the summary of your examination, and it says everything is fine." If the news is not so good, the nurse simply says that the doctor has the results of your examination, and will get back to you as soon as possible.

An alternative approach to improving outcomes is to take a mechanism design perspective. One can model the choice of game form by a "medical establishment" that sets up the form of the interaction between patient and doctor. In this setting, the idea would be to make it "worth the doctor's while" not to reveal anything to a patient of type *L*. In principle this would seem to be feasible. For example one could write a contract in which the patient is told that if his questionnaire answers reveal him to be of type *L* and the doctor reveals anything to him about the operation, then he will receive a massive monetary reward from the doctor. Unfortunately, given the rational fear of frivolous claims, few doctors would work with such a contract.

What of a scheme in which the doctor is fined by the medical authorities for revelation to an *L* type, or rewarded for providing information to an *E* type? There are several practical issues that might make such schemes infeasible. One issue is whether it would be feasible to convince the patient that the doctor's rewards are such that she would be damaged by revealing the truth to the *L* type. A second issue would be the difficulty of defining exactly what information transmission means. What does "Look at this chart. You see, there is **nothing** to worry about" mean? Finally, if the medical authorities are offering monetary rewards to the doctor, would she and some unscrupulous friends be able to create a money machine?

Extensions

Non-Verifiable Doctor Information

If the doctor's information in non-verifiable, then the problem may be the incentive to pass on good news to everyone. Consider the simple model with valenced information of section 6 in which the parameters obey $\gamma > \bar{p} - 0.5$. In this case all equilibria involve complete patient information due to the incentive that the doctor faces to tell the patient the truth in the good state of the world. If the information provided by the doctor is non-verifiable, then this incentive exists even when the true state of the world is bad. In this case, as one would expect, all equilibria involve complete ignorance all around. In equilibrium all patients will be indifferent to what the doctor announces, and the doctor will follow a mixed strategy of information revelation independent of the true state.

The mechanism design perspective suggests that the inefficient outcome may be improved upon by structuring in some financial incentives to overcome the incentive to lie. The doctor could commit to a self-punishment device, such as offering to pay people if she said that the world was a given way and the cumulative evidence indicated otherwise. We remain somewhat skeptical of the utility of these external financial incentives. Reputation effects may be of more help in this case.

There may also be some important additional prizes in the patient's utility function that influence the doctor's decision. What would the doctor say facing a patient who she knew was vulnerable to tremendous ex post feelings of disappointment, in addition to the ex ante anxiety? Lying by pretending that prospects are good may not be such a good idea if the prize in the second period is impacted by the first period belief. In fact, such outcomes appear to show up in the experiment of Morgan et. al., in which the recovery period is impacted by prior information. footnote

The introduction of surprise and disappointment prizes into the utility function is beyond the scope of this paper. It is natural to conjecture that in a full model that allowed for these effects, a far wider array of outcomes would be possible. The doctor may tell the truth to certain types, be uninformative due to a desire to present an optimistic front to some, and equally uninformative due to desire to present a pessimistic front to others. Even in these richer settings, it is very hard to believe that truth-telling would be an equilibrium outcome, unless it was of overwhelming value to the doctor in and of itself.

Non-Verifiable Patient Type and the Incentive to Mislead the Doctor

It is rare for patient information to be verifiable, and it is natural to enquire whether patient self selection in terms of a more arbitrary message space suffices to establish the basic result. The answer is yes, with a caveat. We again assume that the doctor's information is verifiable, and we consider the case with non-valenced information.

At the first stage of the game, allow the patient to select from a general space of messages, $m \in M$. Having made one of these announcements, the patient knows that the doctor's strategy is to tell him that the true state is $\theta = A, B$ with probability $\pi^m(\theta)$. It is no surprise that the situations above in which the two types can be voluntarily induced to reveal have corresponding equilibria in the no-verification game, in which there is purely voluntary self selection into separate messages ("true" messages, in the sense of the revelation principle). All who give message E are true E types, all who give message L are true L types, and the messages are followed by the doctor telling the E types what the state is, and holding back from the L types.

While this natural equilibrium is present in the game with non-verifiable patient types, there are additional equilibria. There exist a continuum of other representative agent equilibria in which all patients use the same probability of giving any message, and the doctor pursues the optimal strategy in the face of the overall population mix of types. Given that there is an even number of type E and type L patients in the population, the doctor can pursue any specific mixed strategy of information revelation. The doctor's mixed strategy must be the same across all patient messages to induce patient indifference. It is clear that these are the only two types of equilibria, since if there is any policy difference on the part of the doctor, then there will be complete separation on the part of the patients.

The situation becomes more subtle if the patient has some reason to hide his type from the doctor. Consider a patient who does not want to hear the information in question, but does not want the doctor to see him as the "fearful" type. Suppose also that this concern is so powerful that it dominates his information preference. He may be tempted to claim to be the "brave" type who can handle the information. In this case, there will be a pooling equilibrium in which all announce that they are the type that would like the information. The doctor tells them all the true state if this is better for the pool, and shows them nothing if this is better for the pool. Given our specific assumption that there are equal number of patients of each type, the doctor is indifferent between informing the pool and not informing the pool.

Other Extensions

One straightforward extension would be to allow for a richer set of differences between patients, including differences in ex ante beliefs, and in the response to disappointment. More broadly, the connection between beliefs and outcomes in medicine goes well beyond the example of the paper, as evidenced by the recent New York Times article on the placebo effect (Blakeslee [1998]).

One important extension to our model would be to consider issues in the period before our game begins. There is some evidence that patients who do not like information choose to delay diagnostic evaluations. footnote If such patterns are widespread, it may already be having very significant impacts on medical costs and medical outcomes. The practical issue would be to design an optimal set of check-up procedures for patients that respects both their preferences for information, and the potential impact on medical costs if they leave a problem undiagnosed for too long.

The most obvious broad generalization of the doctor-patient game is to a broader class of "expert-novice" games. The expert cares about the welfare of the novice, and has valuable information about the external world. The novice knows more about their internal world, including their preferences and their beliefs. In addition to doctor and patient, other expert-novice games include interactions between teacher and student, parent and child, and a psychotherapist and client. As a parent, should you try to make a child aware of such sad realities as the non-existence of Santa Claus as early as possible, or might it be better to wait? Should a teacher tell an over-optimistic student that they are not as good as they appear to think they are, or might there be situations in which it is better to keep their opinion to themselves for a while?

In our current model of how these two parties can interact optimally, we begin with a stage in which the novice transmits some of the information about themselves to the expert. One general question is how the expert can design a questionnaire that allows for optimal self selection by the novices.

In this respect one interesting issue is whether or not to ask direct questions. By simply asking the direct question "are you ready to hear the truth about Santa Claus", the answer escapes. In this case, it may be better to stick with a more abstract and general set of questions in order not to reveal anything that may in and of itself be undesirable knowledge.

The issue gets even more subtle if being asked a particular question induces anxiety in the novice. In this case, it may be best for the expert to take complete control of the situation and makes a decision without asking any questions. A possible preference for lack of control raises intriguing issues for the democratic process. Should the leader lead, or is this always to be regarded as a cheaper but worse option than handing every individual decision back to the people? Is there a role for a constitution that limits the maximum level of passivity of the novice?

The novice is not the only agent who may choose a passive strategy. In many settings, it is easy to imagine that the expert could take a passive role analogous to that discussed in the patient-doctor game. The expert could simply make available a menu of different informational options with as full a description as possible, and then permit the novice to pick up the informational package that they believe best suits them.

We believe that the passive strategy has profound limitations, over and above the already outlined incentive problem for the expert in cases with valenced information. Again, it does not work in the case of Santa Claus, since one label would read "open this document if you wish to know the truth about Santa Claus". The problem is even more complex for the teacher who could pursue the passive strategy by allowing students to self select to move forward when they feel ready. Unfortunately, the student's true readiness to move forward may be more accurately predicted by giving a test on the material. We suspect that the two procedures would show a systematic bias, with many more students claiming to understand than passing the test. In these cases it is the incentive of the students to pretend to be something they are not that presents the problem with the simple strategy of self selection.

Concluding Remarks

We have developed a model of an interaction between a doctor and a patient that has allowed us to explore the doctor's optimal procedure for supplying information to the patient. Our model involves a combination of the psychological approach to decision theory of Caplin and Leahy [1997] with the psychological approach to game theory of Geanakoplos, Pearce, and Stacchetti [1989]. This same combination of approaches may be of value in other applications.

bibitem American Psychiatric Association, 1987, *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev.), Washington, DC.
bibitem Averill, J., and Rosenn, 1972, "Vigilant and Non-Vigilant Coping Strategies and Psychophysiological Stress Reactions During Anticipation of Electric Shock", *Journal of Personality and Social Psychology*, vol. 23, 128-141.
bibitem Baum, A., S. Newman, J. Weinman, R. West, C. McManus (editors), 1997, *Cambridge Handbook of Psychology, Health, and Medicine*, Cambridge, Cambridge University Press.
bibitem Blakeslee, S., 1998, "Placebo Effects Prove So Powerful Even Experts are Surprised", *New York Times, Science Times*, Oct. 13, p.1-4.
bibitem Breznitz, S., 1984, *Cry Wolf*: *The Psychology of False Alarms*, Lawrence Erlbaum Associates, Hillsdale, NJ.
bibitem Brown, L., A. Tomarken, D. Orth, P. Loosen, N. Kalin, and R. Davidson, 1996, "Individual Differences in Repressive-Defensiveness Predict Basal Salivary Cortisol Levels", *Journal of Personality and Social Psychology*, vol. 70, 362-371.
bibitem Cantor, N., and J. Norem, 1989, "Defensive Pessimism and Stress and Coping," *Social Cognition*, vol. 7, 92-112
bibitem Caplin, A., and J. Leahy, 1997, "Psychological Expected Utility Theory and Anticipatory Feelings", Discussion Paper 97-37, C.V.Starr Center, New York University.
bibitem Cornelius, R. and J. Averill, 1980, "The Influence of Various Types of Control on Psychophysiological Stress Reactions", *Journal of Research in Personality*, vol. 14, 503-517.
bibitem Geanakoplos, J., D. Pearce, and E. Stacchetti, 1989, "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, vol. 1, 60-79.
bibitem Janis, I., 1958, *Psychological Stress*, New York, Wiley.
bibitem Jevons, W., 1905, *Essays on Economics*, London, Macmillan.
bibitem Katz, J., H. Weiner, J. Gallagher, and L. Hellman, 1970, "Stress, Distress, and Ego Defenses", *Archives of General Psychiatry*, vol. 23, 131-142.
bibitem Klusman, L., 1975, "Reduction in Pain in Childbirth by the Alleviation of Anxiety During Pregnancy", *Journal of Consulting and Clinical Psychology*, 162-165.
bibitem Kreps, D., 1990, *A Course in Microeconomic Theory*, Princeton, Princeton University Press.
bibitem Kreps, D., and E. Porteus, 1978, "Temporal Resolution of Uncertainty and Dynamic Choice Theory", *Econometrica*, vol.46, 185-200.
bibitem Lazarus, R., 1966, *Psychological Stress and the Coping Process*, McGraw Hill, NY.
bibitem Loewenstein, G., 1987, "Anticipation and the Valuation of Delayed Consumption", *Economic Journal*, vol.97, 666-684.
bibitem Miller, S., and C. Mangan, 1983, "Interacting Effects of Information and Coping Style in Adapting to Gynecologic Stress: Should the Doctor Tell All?", *Journal of Personality and Social Psychology*, vol.45, 223-236.
bibitem Miller, S., 1987, "Monitoring and Blunting: Validation of a Questionnaire to Assess Styles of Information Seeking Under Threat", *Journal of Personality and Social Psychology*, vol.52, 345-353.
bibitem Monat, A., J. Averill, and R. Lazarus, 1972, "Anticipatory Stress and Coping Reactions under Various Conditions of Uncertainty", *Journal of Personality and Social Psychology*, vol.24, 237-253.
bibitem Morgan, J., L. Roufeil, S. Kaushik, and M. Bassett, 1998, "Influence of Coping Style and Precolonoscopy Information on Pain and Anxiety in Colonoscopy", *Gastrointestinal Endoscopy*, vol. 48.

bibitem Roth, S., and L. Cohen, 1986, “Approach, Avoidance, and Coping with Stress”, *American Psychologist*, vol.41, 813-819