

THE SUPPLY OF INFORMATION BY A CONCERNED EXPERT*

Andrew Caplin and John Leahy

How much information should a policy maker pass on to an ill-informed citizen? In this paper, we address this classic question of Crawford and Sobel (1982) in a setting in which beliefs impact utility, as in Kreps and Porteus (1978). We show that this question cannot be answered using a utility function with standard revealed preference foundations. To solve the model, we go beyond the classical model in two respects, relying on the psychological expected utility model of Caplin and Leahy (2001) to capture preferences, and the psychological game model of Geanakoplos *et al.* (1989) to capture strategic interactions.

How much information should a policy maker pass on to a currently ill-informed citizen? This question was first posed formally in the classic sender-receiver game of Crawford and Sobel (1982). In that model, the citizen in question had standard expected utility preferences. In this paper, we enrich the question by allowing for a broader class of preferences, in particular preferences over the timing of resolution of uncertainty.¹ This amendment allows us to address such questions as whether or not a doctor should reveal the truth to a terminally ill patient who is naively optimistic, and whether or not parents should tell their children the truth about Santa Claus. Larger scale policy issues of this form concern the optimal use of warning codes concerning terrorist threats and whether or not bank regulators should immediately pass on new information about systemic credit risk to a worried public.

Given our focus on the resolution of uncertainty, it would be natural to assume that the receiver in our game has Kreps-Porteus preferences (Kreps and Porteus, 1978: henceforth KP). As opposed to the classical expected utility model, these preferences allow individuals to prefer later rather than earlier resolution of uncertainty. The natural modelling procedure would seem to be to append these preferences to a variant of the classical sender-receiver model.

The central point of this paper is that such a mix and match procedure is doomed to failure. The fact that beliefs may impact utility raises a fundamental theoretical issue that cannot be answered using *any* standard model of individual preferences, including the KP model. The issue is simple, yet profound. Standard models of individual preferences, including the KP model, are based on the principle of revealed preference. Yet there are no private choices that can guide the policy maker in our policy problem. In technical terms, revealed preference cannot tell us whether in a particular state s a decision maker prefers to believe

* We thank Nico Frijda, George Loewenstein, and Bob Rosenthal for valuable comments on an early draft, and Kfir Eliaz and David de Meza for equally valuable comments on a later draft. Caplin thanks the C.V. Starr Center at NYU and Leahy thanks the National Science Foundation and the Sloan Foundation for financial support.

¹ To reduce the number of equilibria, we move away from the cheap talk assumption of Crawford and Sobel by assuming that all information is verifiable.

incorrectly that the state is t , or to know the truth. Yet when the sender wants to maximise the utility of the receiver with belief-dependent preferences, knowing this ranking is essential.

In intuitive terms, the insufficiency of the revealed preference approach can be understood readily in the example of the naively optimistic patient. The doctor may know through observation of prior choices that the patient has a general preference for early resolution of information in the KP sense. Yet these choices are mute as to whether or not this same preference would survive *if the news was sure to be bad*, as the doctor knows it to be. The patient may wish to resolve uncertainty but that does not imply a preference for getting specifically bad news early. Decisions on whether or not to leave intact potentially utility-relevant *illusions* cannot be based on preferences revealed through private choice.

In order to make progress with our policy question, we work with a model of individual choice that *does not* have classical foundations. The specific model we use is a version of the psychological expected utility model of Caplin and Leahy (2001) (henceforth PEU). Our model posits a mental 'production function' in which utility relevant mental states such as suspense and anxiety are impacted both by evolving beliefs, and by actual outcomes. Information then affects utility by influencing beliefs, which in turn influence mental states. The policy maker's goal will be to supply information in a manner that maximises the resulting level of private utility.

The particular question of information provision we explore is drawn from the field of behavioural medicine, and concerns the supply of information prior to a stressful medical procedure.² Psychologists have long hypothesised 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. Following the pioneering work of Janis (1958), one of the key questions in the literature concerns the role of information in impacting anxiety. Recent medical research suggests that it is worthwhile to distinguish at least two distinct anxiety-based responses to information during the anticipatory stage of an aversive event (Roth and Cohen, 1986). Some individuals seek to lower anxiety by actively avoiding information at this stage, while others aggressively pursue information to allay anxiety. In terms of the KP model, these two types are characterised by a preference for late and for early resolution of uncertainty respectively. This makes natural the recent experimental finding of Morgan *et al.* (1998) that information avoiders do better if they receive less information, while information seekers do better with more 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 behavioural indices of pain in patients undergoing colonoscopy...

² Behavioural medicine is a vast and rapidly growing area of study, with much focus in particular on stress and anxiety. Many important contributions are summarised in the recent Cambridge Handbook edited by Baum *et al.* (1997).

In addition to informing patients about colonoscopy, a significant goal of precolonoscopy information should be reduction in subjective anxiety.’ (Morgan *et al.*, 1998).

In practical terms, Morgan *et al.* propose use of a survey to sort patients according to their anxiety characteristics. In this paper we set up a simple strategic model that allows us to explore the virtues and the drawbacks of such a questionnaire procedure. In Sections 1 to 4, we plunge directly into the model, using the PEU model of individual preferences. With respect to the structure of the game, we apply the psychological game model of Geanakoplos *et al.* (1989) (henceforth GPS). This framework allows for the payoffs in a game to depend on beliefs about other players’ strategies, thereby accommodating the diverse responses to information allowed for in the PEU model. While there is little that is surprising *per se* in the substantive results, we do point out a credibility problem for the doctor: the ‘no news is bad news’ effect. It is hard for the doctor to hide bad news concerning the likely outcome of the procedure, since the patient will naturally interpret silence in the worst possible light.

In Section 5 we turn to the essential point of methodology, and formally demonstrate that it is not possible to answer our policy question when we restrict ourselves to models of choice based on the principle of revealed preference. When beliefs impact utility, crucial policy questions cannot be answered unless our models are enriched to contain information that cannot be revealed by standard choice experiments. How best to retain intellectual discipline as we develop such models is a crucial question for future research.

1. The Show and Tell Game

1.1. *The Extensive Form*

We model a game between a doctor (‘she’) and a patient (‘he’) who is facing an impending operation. Figure 1 presents the extensive form of this game, which we refer to as the show-and-tell game. The first move is by nature (N), which selects a patient from a population of patients. *Ex ante* the population is evenly divided between two preference types, E and L . We discuss the precise nature of these types below. Once selected, the patient (player 1) chooses whether to show (S) or not show (NS) this type to the doctor.

Nature makes the first move in the tell stage of the game, revealing to the doctor, but not to the patient, whether the operation is of type A or type B . The type of operation determines the distribution of outcomes for the patient. To keep things simple, we assume that there are two possible outcomes, α and β . We allow the precise difference between these outcomes to vary from setting to setting. In a type A operation, α occurs with probability $\bar{p} \in (0.5, 1]$, while in a type B operation the proportions are reversed, with probability \bar{p} of outcome β . All agents enter the game with a common belief that the operation has a 50% chance of being of type A and a 50% chance of being of type B . \bar{p} therefore captures the doctor’s informational advantage over the patient.

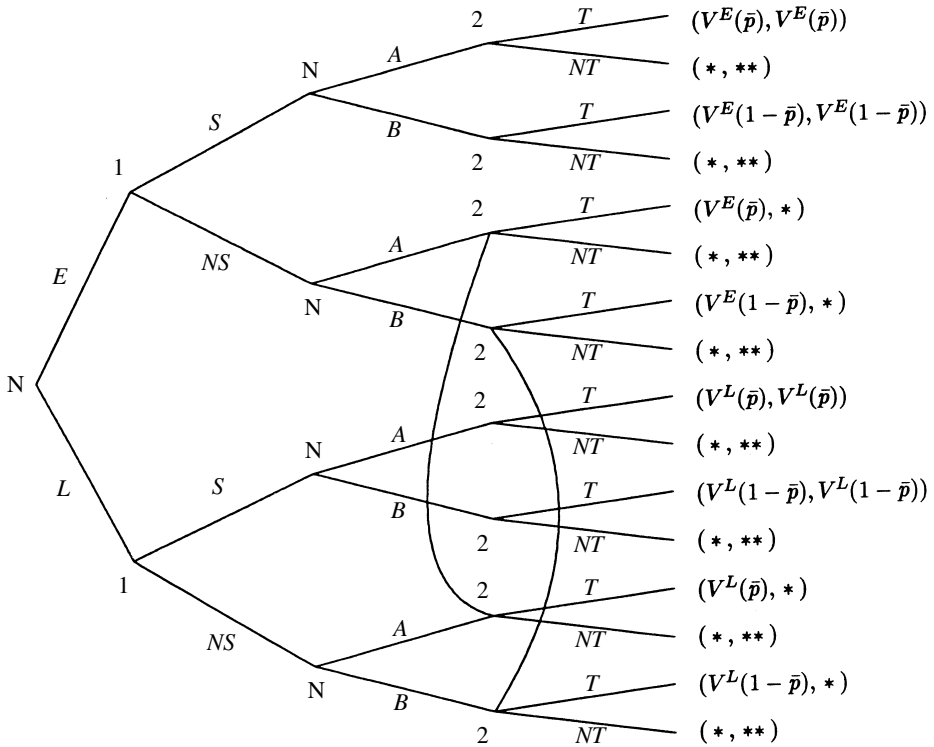


Fig. 1. The Extensive Form Game

Armed with the message from the patient and the nature of the operation, the doctor chooses either to tell (*T*) or not to tell (*NT*) the patient whether the operation is of type *A* or type *B*. Note that the curved lines in Figure 1 connect nodes that belong to the same information set, which correspond to the patient choosing *NS*. The doctor’s action must be the same at each node in these sets.

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. We also assume 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.

An overall player 1 (patient) strategy is a vector $\mathbf{r} = (r^E, r^L) \in [0,1] \times [0,1] = \Sigma_1$ that specifies for each type the probability that he will show the doctor his preference type. We can summarise the strategy of player 2 (the doctor) by the corresponding mixed strategy profile of information revelation. We let π_θ^τ denote the probability that the strategy accords to revelation when the operation is $\theta = A, B$ and the patient communicates $\tau = E, L, NS$, and define $\boldsymbol{\pi}$ to be a complete strategy,

$$\boldsymbol{\pi} = (\pi_A^E, \pi_B^E, \pi_A^L, \pi_B^L, \pi_A^{NS}, \pi_B^{NS}) \in [0, 1]^6 = \Sigma_2.$$

Let $\sigma = (\mathbf{r}, \boldsymbol{\pi})$ denote the overall strategy profile in the game, and let $\Sigma = \Sigma_1 \times \Sigma_2$ denote the set of such strategy profiles.

1.2. Individual Preferences

The doctor is entirely empathetic, deriving utility only from her beliefs concerning patient welfare. With regard to patient preferences, the two distinct types have different attitudes to receiving information about the operation. Type *E* patients prefer learning about the operation as early as possible, while type *L* patients prefer late resolution of uncertainty.

As indicated in the introduction, we capture these preferences by applying a simple version of the PEU model of Caplin and Leahy (2001). This model allows utility to depend on two types of prizes, physical prizes such as consumption, and psychological prizes such as anxiety. The latter can be a function of beliefs about the future.

To apply the model, we conceptualise the patient as experiencing utility at two different stages both occurring after the play of the game has been concluded. The first stage occurs right after the patient has met the doctor but prior to the operation. During this period the patient may experience feelings of anxiety as he anticipates the operation, feelings that may depend on the information that he has received from the doctor. The second stage corresponds to the operation itself. In this period the patient receives one of the physical prizes, α or β , and experiences utility corresponding to this outcome.

We assume that second stage utility, U_2 , is the same for both patient types, and that outcome α is as least as good as outcome β . We normalise by setting $U_2(\alpha) \geq U_2(\beta) = 0$. With respect to the first period, we denote the level of anxiety by $X \geq 0$. All patients are equally averse to anxiety, and we assume that first period utility is linear in the level of anxiety,

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

While there is universal aversion to anxiety, we allow for heterogeneity in its determinants. Both types are reassured when the outcome is more likely to be α but they differ in their reaction to uncertainty. 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.

To be concrete we work inside a linear-quadratic specification. Let p_1 denote the probability that the patient attaches in the first period to outcome α ,

$$X^E(p_1) = -ap_1 - b\left(p_1 - \frac{1}{2}\right)^2;$$

$$X^L(p_1) = -ap_1 + b\left(p_1 - \frac{1}{2}\right)^2.$$

The parameter $a \geq 0$ reflects the possibility that pessimistic beliefs may lower the level of anticipatory utility as suggested by Jevons (1905), while $b > 0$ reflects the impact of pure uncertainty on anxiety. The difference between the patient types is captured in this second term. For a type *L* patient, anxiety is convex in p_1 . This means that the *L* type feels less anxious when uncertain about the outcome of the

operation. Conversely, anxiety is concave in p_1 for the type E patient; other things equal an E type prefers to be certain about the outcome.

Putting the two periods together gives the total utility to the patient of believing that the probability of α is p_1 . We refer to the overall function as the patient's induced expected utility function over beliefs,

$$V^L(p_1) = [U_2(\alpha) + a]p_1 - b\left(p_1 - \frac{1}{2}\right)^2; \quad (1)$$

$$V^E(p_1) = [U_2(\alpha) + a]p_1 + b\left(p_1 - \frac{1}{2}\right)^2. \quad (2)$$

Note that p_1 plays two roles in this payoff function. It directly affects the patient's level of anxiety in the first period and it serves to weight second period outcomes. When the doctor evaluates the patient's utility, she considers these two roles separately. Let p_d denote the probability that the doctor assigns to outcome α after learning the type of the operation. Knowing the type of the operation, the doctor knows that p_d is equal to \bar{p} or $1 - \bar{p}$, each of which may differ from the patient's view p_1 . The doctor's view of the patient's utility is therefore,

$$V_d^\tau(p_1|p_d) = U_2(\alpha)p_d + ap_1 - b\left(p_1 - \frac{1}{2}\right)^2, \quad \tau = E, L.$$

Since the doctor is perfectly empathetic, her utility depends only on her beliefs concerning patient welfare. It remains only to specify how the possible levels of final utility of the two distinct patient types combine to produce empathetic utility for the doctor. We assume that the 'production function' for empathy coincides with the probability weighted average of the two distinct patient utility functions previously specified,

$$qV_d^E(p_1|p_d) + (1 - q)V_d^L(p_1|p_d). \quad (3)$$

2. The Psychological Show-and-Tell Game

2.1. Why Beliefs about Strategies Influence Payoffs

To complete the description of the game, we need to specify payoffs in all branches of the tree. In Figure 1 we have filled in the payoffs in all branches involving revelation by both doctor and patient. 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 remaining payoffs are more subtle. If the patient does not reveal his type, the doctor's payoff depends on her belief about the patient's type, not just the branch of the tree. She must use her beliefs about the strategy that the patient

pursued at the earlier stage of the game to compute her payoff. 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.

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. Instances of two asterisks all involve 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.

2.2. The Psychological Show-and-Tell Game

The defining characteristic of the GPS framework is the role accorded to beliefs about strategies in determining payoffs.^{3,4} The appropriate domain of beliefs is the set of collectively coherent beliefs, $\bar{B} = \bar{B}_1 \times \bar{B}_2$, with generic element $b = (b_1, b_2) \in \bar{B}$.⁵ The set \bar{B}_i , the coherent belief hierarchy of player i , summarises beliefs of all orders about the other player, including beliefs about the other player's strategy, beliefs about the other player's beliefs and so on.

The payoff function to each player in the GPS framework is of the form $u_i(b, \sigma_i, \sigma_{-i})$, with $b = (b_1, b_2) \in \bar{B}$. To show that the show-and-tell game is a psychological game, we have to construct the payoff function associated with a particular pair of strategies and a particular set of collectively coherent beliefs. Given strategies σ and beliefs $b \in \bar{B}$, we can calculate payoffs using our utility specification and Bayes' rule (with one class of exceptions discussed in Section 2.3).

Consider first those branches of the game tree in Figure 1 in which the patient's 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 probability he assigns to non-revelation signifying that the true state is A , which depends 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 $\tilde{\pi}(\bar{b}_1)$. For a patient of revealed type τ , the probability that doctor non-revelation implies

that the true state is A is $\frac{1 - \tilde{\pi}_A^\tau(\bar{b}_1)}{2 - \tilde{\pi}_A^\tau(\bar{b}_1) - \tilde{\pi}_B^\tau(\bar{b}_1)}$. The only situation that this formula

³ See Rabin (1993) for an early application of GPS to a classical game.

⁴ Psychological games differ from Bayesian games in that beliefs enter payoffs directly at the terminal nodes of the tree. In a Bayesian game if one knows the true state of the world, one knows the payoffs; beliefs serve only to provide probabilistic weights to the various states.

⁵ An element $b_i^n \in \bar{B}_i$ represents agent i 's beliefs concerning agent j 's strategy and agent j 's beliefs of order $k < n$. A belief hierarchy is coherent if the marginal distributions of higher order beliefs accord with lower order beliefs.

leaves indeterminate is $\tilde{\pi}_A^r(\bar{b}_1) = \tilde{\pi}_B^r(\bar{b}_1) = 1$, in which case the patient has no initial theory as to why the doctor said nothing. This is one of those exceptional cases discussed in the next subsection.

The branches in which the doctor's utility is marked with a single asterisk involve non-revelation by the patient followed by revelation by the doctor. Here 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, the probability that a non-revealing patient is of type E rather than of type L is $\frac{1 - \tilde{r}^E(\bar{b}_2)}{2 - \tilde{r}^E(\bar{b}_2) - \tilde{r}^L(\bar{b}_2)}$.

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. Again, this exceptional case is taken up below.

The branches in which the doctor's payoff is marked with a double asterisk 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 the patient's 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. It is to such out-of-equilibrium payoffs that we now turn.

2.3. *Out-of-Equilibrium Beliefs and Payoffs*

The unspecified payoffs at this point involve contingencies that are inconsistent with beliefs concerning strategies. Since these contingencies were unreachable according to the supposed strategies, our model does not allow us to pin down payoffs. For example, if the strategy is that all patients reveal their type, we are unable to pin down uniquely the payoff to the doctor either from telling or from not telling these types the true state. Her payoff depends on the assessment she makes in this information set concerning the likelihood that a patient who does not show is of type E . This assessment is technically distinct from the strategy itself, so that we need to pin down assessments as well as strategies before we can compute all payoffs.

To overcome the ambiguity in payoffs in impossible contingencies, our approach will be to fix a specific interpretation for any possible out-of-equilibrium behaviour, and add this interpretation to the data of the model. 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. Fortunately, the precise values of these parameters have no impact on the structure of the analysis that follows.

Why is it necessary for us to add assessments as well as strategies to the data of the model, given that GPS rely on strategies alone? The difference is that GPS treat the belief-dependent payoffs as an exogenous datum of their model. In contrast, payoffs are endogenous in our model: the relationship between beliefs about strategies and final payoffs is itself *derived* from an underlying economic model. Whenever payoffs are endogenous and extensive form reasoning is required, the ambiguity we noted above is likely to arise and a strategy such as adding assessments to the data of the model will become necessary. The fact that this point has not been noted before stems from the very limited use of the GPS model in applications.

2.4. Psychological Equilibrium

In the GPS framework, an extensive form game tree, beliefs about nature's moves, and utility functions $u_i(b, \sigma_{-i})$ determine a specific extensive psychological game, Γ . Given any coherent belief system $\bar{b} = (\bar{b}_1, \bar{b}_2) \in \bar{B}$, $\Gamma(\bar{b})$ denotes the standard extensive form game with final payoffs corresponding to beliefs \bar{b} .

For a strategy profile $\sigma \in \Sigma$ to be a psychological Nash equilibrium of Γ , it must be optimal for each player to pursue their corresponding strategy σ_i in a setting in which the payoffs are determined using common knowledge that the strategy is indeed σ . Thus beliefs $b \in \bar{B}$ in which it is common knowledge that a particular strategy profile $\sigma \in \Sigma$ is being played play a special role in the theory. We denote this special class of beliefs as $\beta(\sigma) \in \bar{B}$. A pair $(\hat{b}, \hat{\sigma}) \in \bar{B} \times \Sigma$ is a psychological Nash equilibrium (PNE) of the psychological game G if and only if (i) $\hat{b} = \beta(\hat{\sigma})$; and (ii) for each player i , and for each $\sigma_i \in \Sigma_i$

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

Given that we are looking at an extensive form game, we are interested in refinements. GPS focus on a specific refinement of PNE: sequential psychological equilibrium (SPE). To define these equilibria, let $\mu \in M$ be a belief system that associates with each information set a probability distribution over the nodes of that set. An SPE is a triple (b, σ, μ) that satisfies the following properties: (i) (b, σ) is a PNE, and (ii) (σ, μ) is a sequential equilibrium of $\Gamma(b)$. The set of sequential psychological equilibria of the game Γ is denoted $S(\Gamma)$.

To apply these definitions to our model, we need to make one simple amendment. As pointed out above, our psychological game is only defined when we include the parameters $\lambda = (\lambda_A, \lambda_E) \in [0, 1]^2$ defining possible out-of-equilibrium beliefs. Adding this to the data of the problem changes none of the definitions that follow. In particular, we will be interested in looking at the SPE of the resulting psychological games. However, it is necessary for us to run one additional consistency check that is not called for in the original GPS model with exogenously given payoffs. Specifically, we want the assessment μ to be consistent with the parameter λ . This assumption has bite only in the situation in which player 1 is

certain to show according to the strategies. In this situation, it would be bizarre to allow the doctor to have any beliefs other than those specified by λ_E , even if these assessments were possible in a sequential equilibrium of the resulting game. In essence, an equilibrium must involve a fixed point not only in strategies, but also in assessments.

3. Neutral Information and Efficiency

We solve for the SPE for the special case of our model in which the two pure prizes are indifferent, $U_2(\alpha) = U_2(\beta) = 0$. In this case, only the degree of uncertainty concerning the outcome matters, not the outcome itself. We call this the case of *neutral information* as opposed to the case of valenced information that we analyse later. It simplifies matters without affecting the results to assume that $b = 1$. This allows us to simplify the induced expected utility functions of (1) and (2),

$$V^L(p_1) = -\left(p_1 - \frac{1}{2}\right)^2;$$

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

To gain insight into the solution, we consider the extensive form of the game $\Gamma(\mathbf{r}, \boldsymbol{\pi})$ with $\mathbf{r} = (\frac{1}{3}, \frac{2}{3})$ and $\boldsymbol{\pi} = (\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3})$. In this game the E type reveals with probability $1/3$, the L type with probability $2/3$, and the doctor reveals the true state with probability $1/3$ to those who announce E , with probability $2/3$ to those who announce L , and with probability $1/3$ and $2/3$ in states A and B respectively to those who choose not to reveal their type, NS .

To understand the payoffs, consider first the doctor's payoff if the patient does not reveal type. Given the patient strategy $\mathbf{r} = (\frac{1}{3}, \frac{2}{3})$, 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 $V^E(1/2)$. Similarly

if she does not tell a patient of revealed type L the state, then she knows that the patient will derive utility $V^L(1/2)$.

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 $2/3$ probability that the agent is of type E , while the doctor's non-revelation strategy leaves the patient believing that there is a $2/3$ probability that the true state is A . The doctor thus derives empathic expected utility of amount Ψ ,

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

To check whether the strategies $\mathbf{r} = (\frac{1}{3}, \frac{2}{3})$ and $\boldsymbol{\pi} = (\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3})$ constitute a sequential psychological equilibrium, we look at the game with the payoffs calculated in the above manner as a standard game of incomplete information, and consider first sequential rationality of the proposed strategies. Given these payoffs, if the patient reveals that he is an E type, it is a dominant strategy for the doctor to reveal, since her payoff from revelation is $V^E(\bar{p}) = V^E(1 - \bar{p}) > V^E(1/2)$. It is a dominant strategy for her not to reveal to the L type, since her payoff from revelation is $V^L(\bar{p}) = V^L(1 - \bar{p}) < V^L(1/2)$. With respect to the group NS , the doctor's dominant strategy is to tell the patient the truth, since $\frac{2}{3} V^E(\bar{p}) + \frac{1}{3} V^L(\bar{p}) > \Psi$. Given this, the E types will be 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. Since the strategies $\mathbf{r} = (\frac{1}{3}, \frac{2}{3})$ and $\boldsymbol{\pi} = (\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3})$ are not sequentially rational in this game, we conclude that they do not constitute a sequential psychological equilibrium.

Proposition 1 establishes the essential properties of the outcomes associated with the equilibrium set $S(\Gamma)$, which turn out to be independent of the values of the parameters λ_A and λ_E defining out of equilibrium payoffs.

PROPOSITION 1 *There are SPE of Γ having the property that all L patients are given no information, and all E patients are provided with complete information. In 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 . There are also fully revealing equilibria in which both E and L types learn the truth for sure. There are no other types of equilibria.*

Proof. The precise nature of the equilibria depends on the value of the parameter $\lambda_E \in [0,1]$ that specifies common beliefs about the patient type of a no-show in a situation in which all are expected to show. To simplify, we assume for now that $\lambda_E = 0$, so that such a type is known to be in favour of late resolution, relaxing this assumption later. In this case, we first establish that the most obvious combination (b^*, σ^*, μ^*) of beliefs, strategies, and assessments constitutes an SPE. The strategy σ^* is for the patient to show for sure, the doctor to tell the truth for sure to an E type, and to suppress it for sure from an L type and from a no-show. We set $b^* = \beta(\sigma^*)$. Finally, the assessment μ^* is that a no show is of type L for sure.

It is straightforward to confirm that this triple constitutes an SPE. It is certainly optimal for the doctor to leave the L type ignorant in both states, since this gives these types maximised utility $V^L(1/2)$. Given her assessment, the same holds true for the no shows. For the E types, who are expecting to learn the truth for sure, the doctor's payoff to deviating to non-revelation depends on the parameter $\lambda_A \in [0,1]$ denoting 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 but it was not. Regardless of the value of this parameter, the payoff can be no higher than that associated with revelation, so that again the doctor's strategy is optimal. Turning now to the patient's strategy of revelation, note that it too is optimal, since there is nothing better for the late type than certain ignorance, and for the early type than certain knowledge, which can be assured by testing. With sequential rationality established, it remains only to confirm consistency. This is immediate, using totally mixed strategies in which the probability of a late resolver deviating to non-revelation shrinks one order less quickly than does the corresponding probability for an early resolver.

The SPE above has the precise structure asserted in the Proposition. We now identify a fully-revealing equilibrium. In one such equilibrium, all patients shows with probability $1/2$, while the doctor reveals the truth to all types in state A , and suppresses it for sure in state B . The doctor believes no shows to be of either type with equal probability. To see that this is an equilibrium, note that on hearing nothing, a patient assumes that the state is B for sure. The fact that non-revelation is interpreted as implying that the state is certain to be B implies that the doctor has no reason to deviate from revelation in state A even with a declared L type. In fact, with these strategies in place, the corresponding psychological game involves both parties being certain of the outcome (possibly incorrectly!) regardless of what the doctor chooses. Hence all strategies are sequentially rational. Consistency is immediate.

It remains to show that there are no other types of equilibria. First we show that in any SPE, E types are informed for sure in at least one state. If to the contrary there was a strictly positive probability of their being uninformed in both states, then the silence would lead to confusion, and the payoffs engendered by these beliefs would make it optimal to supply information for sure in both states. Given that E types learn the truth for sure in one state, they are always fully informed in any SPE.

To complete the proof, it remains only to show that L types cannot be partially informed in equilibrium: they are either fully informed or completely ignorant. For a declared L type to be partially informed requires him to be informed with probability strictly below 1 in both states. Yet in this case, silence would lead to confusion, and the payoffs engendered by these beliefs would make it optimal to suppress information for sure in both states. The only remaining possibility is for an L type to not declare, join group NS , and thereafter be partially informed in equilibrium. Yet we know that this is not possible, since in a purported equilibrium of this type, E types would rationally reveal themselves, leaving only L types in group NS , enabling us to reiterate the argument just given for L types.

It remains only to confirm that the same structural results concerning the SPE go through for values of the parameter $\lambda_E \in (0,1]$. To see this, note first that the initial SPE is little changed by this alteration. The strategy σ^* is for the patient to show for sure, the doctor to tell the truth for sure to an E type, to suppress it for sure from an L type, and to adopt the strategy optimal for the mix associated with λ_E for no-shows. We set $b^* = \beta(\sigma^*)$. The assessment μ^* is that a no show if of type E with probability λ_E . Nothing changes in the argument establishing that this is indeed an SPE. Note also that the fully revealing equilibrium is independent of λ_E . Finally, note that the proof that E types are fully informed and that L types can never be partially informed in any equilibrium also go through just as before. ■

Note that the doctor is happy with all but the fully revealing equilibria in this game, since these other equilibria achieve her first best expected utility,

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

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 existence of efficient equilibria, and their achievability in 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.

4. Valenced Information and Inefficiency

4.1. Valenced Information and 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, $U_2(\alpha) = 1$ and $U_2(\beta) = 0$. To keep the analysis simple, we set $b = 1$ in (1) and (2),

$$V^L(p_1) = (1 + a)p_1 - \left(p_1 - \frac{1}{2}\right)^2;$$

$$V^E(p_1) = (1 + a)p_1 + \left(p_1 - \frac{1}{2}\right)^2;$$

with $a \geq 0$.

4.2. 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 $V^E(\bar{p})$. 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, just as with the game with non-valenced information, this game may have equilibria with information about the operation revealed to those who show that they are of type E , and held back from those of type L . 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 . Given the proposed strategies, if the doctor does not tell the patient the state, the doctor gets payoff $V_d^L(\frac{1}{2}|\bar{p})$. If she does, she gets payoff $V_d^L(\bar{p}|\bar{p})$. It is rational for her not to reveal the information if and only if,

$$V_d^L(\bar{p}|\bar{p}) = (1 + a)\bar{p} - \left(\bar{p} - \frac{1}{2}\right)^2 \leq \bar{p} + \frac{a}{2} = V_d^L\left(\frac{1}{2}|\bar{p}\right) \Leftrightarrow a \leq \left(\bar{p} - \frac{1}{2}\right).$$

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 $a > (\bar{p} - \frac{1}{2})$ the doctor knows that even the type L patient's utility would 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, $a > (\bar{p} - \frac{1}{2})$, ensures that if the operation is of type B , then the doctor knows that the E type does not want to know. Of course, 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 , silence implies that the true state of the world is B .

Overall it is natural to conjecture that in cases with $a > (\bar{p} - \frac{1}{2})$, all equilibria to the game will involve both types of patient knowing the true state for sure. This is confirmed in Proposition 2.

PROPOSITION 2 *With $a > (\bar{p} - \frac{1}{2})$, all SPE of Γ^{VA} are fully revealing. 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 .*

Proof. We first establish that the existence of a fully revealing SPE (b^*, σ^*, μ^*) . In one such equilibrium, all patients show with probability $1/2$, while the doctor reveals the truth to all types in state A , and suppresses it for sure in state B . The doctor believes no shows to be of either type with equal probability. To see that this is an equilibrium, note that with the payoffs associated with these beliefs, all types are made better off if the doctor reveals the truth in state A , since hearing nothing will convince them that the state is B for sure. Given this, the doctor's strategy in state B has no impact on payoffs, since the assumption will be made that the state is

B for sure regardless of whether or not she reveals this information. The proof of sequential rationality is completed by noting that both *E* and *L* types get precisely the same payoffs regardless of whether or not they reveal their type. As usual, consistency is immediate.

It remains to show that there are no other types of equilibria. The key here is that, regardless of the assumed strategy, the optimal strategy for the doctor is to reveal state *A* to anyone regardless of their type. The only possible exception to this occurs when non-revelation is taken anyway to imply that the state is *A*. In order for this to be true, it must be that the doctor reveals for sure in state *B*, which is itself inconsistent with the underlying patient preference to remain hopeful rather than completely pessimistic. Given that all equilibria involve revelation for sure to all types in state *A*, it is immediate that any such equilibria are fully revealing: either state *B* is revealed directly, or it is inferred from the doctor's silence. ■

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.

4.3. *Changing the Game*

One attempt to provide a commitment mechanism would be for the doctor to follow the passive strategy alluded to in Section 3, 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?

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?

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. Possibly the best hope lies in an effort to automate the provision of information so that it is handled by a machine rather than a human being. In some contexts, it may be possible to have information delivered by an effectively neutral

'mechanism designer'. In this spirit, Caplin and Eliasz (2003) propose a mechanism for reducing the spread of AIDS by reducing the anxiety associated with testing for the disease. Again, the mechanism can work only if test results are automated, and information kept out of the hands of empathic caregivers.

5. PEU, KP and Revealed Preference

While we have solved our model in an apparently natural manner, we have done so using a very non-standard model of individual preferences. In particular, the PEU model is based on hypotheses concerning the determinants of anxiety, a mental state. In classical choice theory by contrast, it is standard operating procedure to apply the principle of revealed preference and to limit the domain of the utility function to the set of privately available choices. According to this principle, it is not necessary to discuss mental states but rather their projection down onto the space of private choices. In this Section, we show that our model would have been incoherent had we restricted ourselves to using models based on revealed preference. Non-standard preferences are absolutely required to model the issue at hand.

In the spirit of revealed preference, a private agent can determine how accurate a signal to observe concerning the true state of the world. In order to model this choice, we follow the lead of Kreps and Porteus, constructing first the space of temporal lotteries. To construct this space, we begin with the space of physical outcomes, z . In our case the only physical prizes are the second period outcomes to the operation, $z = \{\alpha, \beta\}$. We then define L_1 as the set of all possible period 1 (post-game, pre-outcome) beliefs concerning the outcome of the operation,

$$L_1 = \{p_1 \in [0, 1] \text{ such that } p_1 = \Pr_1\{z = \alpha\}\}.$$

The space of temporal lotteries, L_0 , is the class of distribution functions on these probabilities. These are all the possible private choices concerning how strong a signal to seek today concerning the true outcome of the operation. It is only choices in this domain that are potentially observable by direct choice. Hence a classical model of preferences cannot have a domain any richer than this.

To force our model to fit with the principle of revealed preference, we project it down onto the domain of temporal lotteries. What we are left with as the revealed preference counterpart to the PEU model is the induced expected utility function over beliefs of (1) and (2). In the special case with $U_2(\alpha) = 1$ these equations read,

$$V^L(p_1) = (1 + a)p_1 - b\left(p_1 - \frac{1}{2}\right)^2; \quad (4)$$

$$V^E(p_1) = (1 + a)p_1 + b\left(p_1 - \frac{1}{2}\right)^2 \quad (5)$$

with $a, b \geq 0$. Given that these are expected utility functions, temporal lotteries are ranked according to the corresponding expected value of these functions. Given $F, G \in L_0$,

$$F \succeq G \quad \text{if and only if} \quad E_F[V^\theta(p_1)] \geq E_G[V^\theta(p_1)], \quad \theta = E, L.$$

The above model of choice is closely related to the Kreps-Porteus model of preferences over the timing of the resolution of uncertainty. The KP axioms are equivalent to the existence of a strictly monotone function $K: [0,1] \rightarrow R$, such that given $F, G \in L_0$,

$$F \succeq G \quad \text{if and only if} \quad E_F[K(p_1)] \geq E_G[K(p_1)].$$

It is immediate that the PEU model satisfies the KP axioms provided $1+a \geq b$ (to ensure monotonicity). In such cases, the KP model is the revealed preference counterpart to our PEU function. We now explore the implications of taking the KP model of (4) and (5) as defining of private preferences.

The critical observation is that if we define the model only by (4) and (5), it no longer has a well-defined solution. To see this, consider first a full PEU model as analysed in the last Section with $a = 0$ and $b = 1$. According to Proposition 1, this model has an equilibrium with information suppression from a type L . By way of contrast, a PEU model with $a = 2$ and $b = 3$ has only fully revealing equilibria. Yet these two distinct parameter settings give rise to identical KP utility functions. Substitution of $a = 0$ and $b = 1$ in (4) yields,

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

Substitution of $a = 2$ and $b = 3$ yields,

$$V^L(p_1) = 3p_1 - 3\left(p_1 - \frac{1}{2}\right)^2.$$

In terms of the KP model, these two different PEU models are identical. In both cases, the subject rejects all information that is privately available. Apparently, knowing that information is rejected is insufficient to produce a well-defined model, since one and the same KP model of choice is consistent with entirely different solutions.

In terms of the PEU model, it is straightforward to understand the difference between the case in which $a = 0$ and $b = 1$ and the case in which $a = 2$ and $b = 3$. The former case involves no anticipatory pleasure at the better outcome. The only factor impacting preferences is the level of anxiety, which is higher as the outcome becomes more certain. In this case, even good news is aversive and the doctor will withhold such news from the patient. In contrast, in the case with $a = 2$ and $b = 3$, optimistic beliefs more than compensate for any additional anxiety and the doctor prefers to pass on good news to the patient.

The specific problem noted above is not merely a small quirk of the data: it is absolutely fundamental to the problem of information transmission. In common sense terms, a private decision to avoid information may arise from many different psychological forces. For example, it may result from a love of suspense, or alternatively, from fear of hearing bad news. Obviously, these distinct motivations have opposite implications for a policy maker deciding whether or not to pass on good

news. One should never provide information to an individual who loves suspense, while good news is beneficial to one who is afraid of bad news. We all know this in our bones. When making empathic decisions on what to communicate, we routinely base these on hypotheses concerning motivation, not merely a history of privately observed choices. The formalities above merely confirm this piece of obvious common sense.⁶

There is another more formal way to see the limitation inherent in using a utility function with a revealed preference foundation to try to answer our problem of information transmission. Whatever utility function is used, it will have to assign a private utility to two distinct end nodes: one in which the actual outcome was known by the patient ahead of time, and one in which it was not. There is simply no private choice that enables us to identify the impact of this difference in *ex ante* beliefs on private welfare, given that the actual outcome is not subject to variation.

The analysis above shows that the classical revealed preference approach to the theory of choice is insufficient to answer a potentially important class of policy questions in which the policy maker must decide how much information to share with private agents. We believe that this negative result will apply not only to the specific model above but to a broader class of non-expected utility models. Hence we see our point of method as presenting a challenge, not only to those interested in the design of health information policies, but to all choice theorists interested in policy. In particular, what discipline can be used to augment the classical principle of revealed preference when this principle is no longer adequate? In theoretical terms, the question concerns the appropriate domain for the theory: over what class of objects should preferences be defined? Certainly, we believe that this domain has to be broad enough to ensure that answers are available for the richest possible class of policy questions but is there anything more precise that can be said? Should the domain be kept to the minimum consistent with a specific class of policy questions, or should it be less restricted?

Once one has expanded the domain appropriately, there is another challenge. How can we go about gathering evidence on utility in realms in which observing private choice is not an option? We believe that observation of empathic choices, surveys, physiological measures and possibly even brain scans, will increasingly be seen as valid sources of additional evidence. It is with these additional sources of evidence in mind that we formulated our own PEU, used above. As techniques for measuring emotions such as fear and suspense improve, we believe that a model such as ours will be seen as increasingly natural.

⁶ This result suggests that far more thought needs to be devoted to the interaction between beliefs and preferences. Indeed a rich investigation of this interaction may call into question the universal applicability of Bayes' rule, which we lean on heavily in our analysis. In particular, Eliaz and Spiegel (2003) demonstrate that it may be necessary to develop non-Bayesian models to rationalise many 'stylised facts' concerning information acquisition and its consequences.

6. Concluding Remarks

This paper makes both a substantive and a methodological contribution. On the substantive side, we explore the optimal procedure for supplying information to a potentially anxious subject. On the methodological side, we outline a class of policy questions that can be answered only by going beyond the revealed preference approach to choice behaviour.

New York University.

Date of receipt of first submission: December 2001

Date of receipt of final typescript: September 2003

References

- Baum, A., Newman, S., Weinman, J., West, R. and McManus, C. (eds.) (1997). *Cambridge Handbook of Psychology, Health, and Medicine*. Cambridge: Cambridge University Press.
- Caplin, A. and Eliaz, K. (2003). 'AIDS policy and psychology: a mechanism design approach', *Rand Journal of Economics*, vol. 34, pp. 631–46.
- Caplin, A. and Leahy, J. (2001). 'Psychological expected utility theory and anticipatory feelings', *Quarterly Journal of Economics*, vol. 116, pp. 55–79.
- Crawford, V. and Sobel, J. (1982). 'Strategic information transmission', *Econometrica*, vol. 50, pp. 1431–51.
- Eliaz, K. and Spiegel, R. (2003). 'Anticipatory feelings and choices of information sources', New York University Working Paper.
- Geanakoplos, J., Pearce, D. and Stacchetti, E. (1989). 'Psychological games and sequential rationality', *Games and Economic Behavior*, vol. 1, pp. 60–79.
- Janis, I. (1958). *Psychological Stress*, New York: Wiley.
- Jevons, W. (1905). *Essays on Economics*, London: Macmillan.
- Kreps, D. and Porteus, E. (1978). 'Temporal resolution of uncertainty and dynamic choice theory', *Econometrica*, vol. 46, pp. 185–200.
- Morgan, J., Roufeil, L., Kaushik, S. and Bassett, M. (1998). 'Influence of coping style and pre-colonoscopy information on pain and anxiety in colonoscopy', *Gastrointestinal Endoscopy*, vol. 48, pp. 119–27.
- Rabin, M. (1993). 'Incorporating fairness into game theory', *American Economic Review*, vol. 83, pp. 1281–302.
- Roth, S. and Cohen, L. (1986). 'Approach, avoidance, and coping with stress', *American Psychologist*, vol. 41, pp. 813–9.