RAND Journal of Economics
Vol. 34, No. 4, Winter 2003
pp. 631–646

AIDS policy and psychology: a mechanism-design approach

Andrew Caplin*

and

Kfir Eliaz*

Economic theorists have given little attention to health-related externalities, such as those involved in the spread of AIDS. One reason for this is the critical role played by psychological factors, such as fear of testing, in the continued spread of the disease. We develop a model of AIDS transmission that acknowledges this form of fear. In this context we design a mechanism that not only encourages testing but also slows the spread of the disease through voluntary transmission. Our larger agenda is to demonstrate the power of psychological incentives in the public health arena.

1. Introduction

■ Economic theorists have devoted great effort to designing mechanisms to reduce the damage caused by externalities. How best to slow the spread of AIDS would seem to be an important case in point. Yet despite the pioneering efforts of Philipson and Posner (1995) and Kremer (1996), economic theorists have largely ignored this question. Indeed they have given little attention to *any* health-related externalities, despite their profound social importance.

One factor that differentiates health-related from standard externalities is the central role played by psychological factors. This critical role of psychology is hinted at by Philipson and Posner when they discuss the role of *fear* in limiting the efficacy of certain AIDS policies. In particular, they discuss the potential impact of verifiable "AIDS-cards" that offer proof to all that one does not have the virus. In an idealized version of this scheme, they speculate that there would be assortative matching, with those who were verified to be clear of the disease matching only with others of their type. Yet, following the empirical findings of Lyter et al. (1987), they argue that not many would be willing to take such a test for psychological reasons:[1]

* New York University; andrew.caplin@nyu.edu, kfir.eliaz@nyu.edu.

We thank two anonymous referees for helping us to substantially improve the article. We also thank George Akerlof, Jay Bhattacharya, Jim Burgess, Marty Gaynor, Bart Hamilton, Botond Koszegi, Alessandro Lizzeri, John Moran, Tomas Philipson, Matthew Rabin, Ronny Razin, Ran Spiegler, Andrew Schotter, Jeroen Swinkels, and Leeat Yariv for their valuable comments. In addition, we thank the Management Science Group at the Veterans Administration for sponsoring the 13th Annual Health Economics Conference. Finally, we thank the C.V. Starr Center at NYU for financial support.

[1] Lyter et al. find that many patients are reluctant to learn their HIV status even when confidentiality is guaranteed. Among the most common reasons cited for rejecting this information is precisely the anticipation of severe psychological distress if the result were to be positive.

many people are fearful of tests which may show they are doomed even if the probability of that result is very low (Philipson and Posner, 1995, p. 472).

In our view, psychological realities of this type need no longer be seen as barriers to progress in economic theory. Rather, they are profoundly enriching. The time has come not only to acknowledge their importance, but also to incorporate them into policy analysis. In this spirit we present a theoretical approach to AIDS policy that explicitly incorporates fear of the form suggested by Philipson and Posner. We use our approach to reassess the potential for certification policies to reduce the spread of the disease. We outline circumstances in which a variant of the AIDS-card scheme may be a very effective policy tool, even when fear is profound.

We begin our analysis in Section 2 by developing a strategic model of the spread of AIDS, and the potential role of certification in limiting its spread. We confirm the conjecture of Philipson and Posner that in the absence of fear, there is an equilibrium in which all agents test, and matching is assortative. In Section 3 we incorporate into the model a fear-induced preference for late resolution of uncertainty and confirm that it may indeed render the certification policy ineffective. If fear is sufficiently important, then not everyone tests, and the disease continues to be spread by those who are HIV-positive.

What can a policy maker do to reduce the spread of AIDS? Section 4 lays out our vision of the feasible set of policies. We assume that the policy maker is able both to assess the health status of private agents and to pass certifiable messages back to them based on the results of these assessments. We analyze two different classes of policy: conditional mechanisms that condition the procedure of sending messages on the test results of all individuals, and unconditional mechanisms that have less flexibility in their choice of messages. Throughout, we assume that the policy maker has no direct control of any subsequent social interactions among the private agents. To affect change, the policy maker must not only induce voluntary testing and certification, but also structure the certification process so that it changes subsequent patterns of sexual behavior.[2]

Section 5 explores the details of policy design in an important special case. We consider a policy maker whose fundamental goal is to stop the spread of AIDS. We provide conditions under which certification policies can be used to establish an equilibrium with absolutely *no* new infections. This provides a positive answer to the key question of feasibility. Under the assumed conditions, which may include the presence of a high level of fear, policies exist that can stop cold the spread of infection.

In addition to establishing the value of psychological incentives, Section 5 also clarifies the distinction between the conditional and unconditional mechanisms. Unconditional mechanisms are blunt tools. The only way that they can be used to stop the spread of AIDS involves simultaneously blocking many *ex post* advantageous trades by inducing doubt concerning the HIV status of potential partners. Conditional mechanisms are considerably more delicate, and they enable the policy maker to not only stop the spread of AIDS but also achieve the secondary goal of allowing as many risk-free matches as possible to take place, while keeping fear to a minimum.

While conditional mechanisms have clear advantages over unconditional mechanisms, they also have practical disadvantages. One disadvantage is that they are far more complex, since the test result of any one individual depends on the results of others. This would make them far more difficult to mechanize than the unconditional mechanisms. If the tests cannot be mechanized, it is likely that the policy would have to be left in the hands of individual physicians. This raises the question of credibility. We show in Section 6 that a caring physician with a healthy patient would not agree to pass on the ambiguous message called for by our mechanisms.[3] In practical terms, our unconditional mechanism may be more readily implemented than our conditional mechanism.

Although the details are somewhat intricate, the fundamental point of our analysis is simple.

[2] Given our fundamental interest in psychological incentives, we do not consider variations in monetary incentives of the type analyzed by Philipson and Posner.

[3] Quite apart from the issue of credibility are the issues of morality and of legality. Ruling out distortions as immoral or illegal severely restricts the class of mechanisms that are available. In any case, our results suggest that these rules may be problematic if one is a strict humanist.

There are strong health-based incentives to test for AIDS, but fear may override these incentives. Our resolution of the problem is to decrease the informativeness of a bad test result, mitigating the fear of bad news and thereby allowing the health-based incentives to reassert their primacy. A similar approach to policy may be of value in the many other medical settings in which fear-based avoidance behavior is believed to play a role. For example, according to Dr. Timothy Johnson, ABCNEWS's medical editor (see Cohen, 2002), "Study after study has shown that men are more reluctant to face up to worrisome symptoms or go to the doctor for checkups. And that is probably one big reason why men's life expectancy, which in the early 1900s was virtually the same for both sexes, now lags behind by approximately six years."[4] In addition, there may be other policy options worthy of exploration, such as using "fear appeals" to induce a private desire for knowledge (Witte, 1998; Witte and Allen, 2000; and Caplin, 2003). As research on these subjects develops, we believe that "behavioral epidemiology" will take its place alongside economic epidemiology (pioneered by Philipson (2000) and others) as a guide to policy makers trying to influence health outcomes.

The theoretical analysis that follows has a certain amount of novelty. To the best of our knowledge, ours is the first example in which the Kreps and Porteus (1978) model of preferences over the timing of the resolution of uncertainty has been placed into a mechanism-design framework. This combination calls for use of the psychological game apparatus of Geanakoplos, Pearce, and Stacchetti (1989). Yet our article represents only the smallest of first steps in the larger program of incorporating psychological phenomena into policy analysis. Economic theory itself is in need of substantial expansion if we are to incorporate a sophisticated understanding of psychological phenomena into our analysis.

2. The basic model

■ We model the social context underlying the spread of AIDS as an extensive game with imperfect information but perfect recall (see Osborne and Rubinstein, 1994). We denote this game by Γ.

□ **The extensive-game form.** Society consists of a fixed finite set of individuals.[5] We focus in particular on the three-player case, since this is the smallest number in which competitive forces can come into play to induce testing. In Section 6 we comment on how the model can be extended to allow for a large population of players. The game itself is played in four stages as follows:

(i) *Determination of players' types*. Each player has probability p of being infected and the probabilities are independently distributed across the three individuals. We let $t = (t_1, t_2, t_3)$ denote the vector of types, with $t_i \in \{(+), (-)\}$, where $(+)$ stands for "infected" (HIV positive) and $(-)$ stands for "healthy" (HIV negative). Players are ignorant of their types, but the probability distribution according to which their types is determined is common knowledge.

(ii) *Private testing decisions*. At the second stage of the game, all three players simultaneously decide whether or not to test for AIDS. We let a^0 denote the vector of testing decisions, with $a_i^0 \in \{T, NT\}$, where T represents a decision to test and NT represents a decision not to test. Each player observes only his own action and is left to infer the decisions of others.

(iii) *Message transmission*. Following the testing stage, nature moves by sending a message m_i to each player i. Messages are sent simultaneously to all players and each player observes only his own message. In the current context, there are only two types of messages: a certificate of health C and no message (or the "null message"), denoted $\emptyset$.

[4] There is also strong evidence that many are too fearful to wish to see the results of genetic tests that may potentially contain very bad news about future health (Caplin and Leahy, 2001).

[5] For simplicity, all players in the game are male, while the policy maker is female.

We assume that a player receives a certificate if and only if he tested and was found to be healthy; otherwise, he receives nothing. We also assume that the set of certified individuals, denoted $N^C \subset \{1, 2, 3\}$, is publicly known.[6]

(iv) *Matching stage*. All players simultaneously decide on a matching strategy, denoted a^1. There are two possible matching strategies: strategy C, which represents a commitment to match only with a certified player, and strategy NC, which represents a commitment to match with any player, certified or not.[7] If there exists a pair of players such that each satisfies the matching conditions of the other (e.g., two certified players who announce C or a certified player who announces NC and a noncertified player who announces C), then those two players match. A pair of such players is said to be eligible to match. If there exist more than one such pair, then each eligible pair has an equal chance of being matched. In all other cases, no matching occurs.

The above matching process can be interpreted as the reduced form of the following dynamic process. First, a pair of players is randomly drawn. If this pair is eligible to match, then a match occurs and the game ends. Otherwise, the two players return to the pool and a different pair is drawn. If this second pair is eligible to match, a match occurs and the game ends. Otherwise, a third distinct pair is drawn and either a match takes place or the game ends with no matches. Note that the strategy imposes a form of anonymity, in that proposals are made to player types rather than to individual players. This captures the idea that these players are not intimately acquainted at the point of matching.

□ **Outcomes.** Let Z be the set of terminal histories in Γ. Each terminal history $z \in Z$ is associated with a pair (s, h). Here s records any match that has taken place and is either a pair of players or the null set,

$$s \in \{\{1, 2\}, \{2, 3\}, \{1, 3\}\} \cup \{\emptyset\},$$

while $(h_i) \in \{(+), (-)\}$, $i = 1, 2, 3$, is the final health states of the players, which depends on s and t. If $i \notin s$, then $h_i(s, t) = t_i$; if $t_i = (+)$, then $h_i(s, t) = (+)$. Finally, suppose $s = \{i, j\}$. If $t_j = (+)$, then $h_i(s, t) = (+)$, otherwise $h_i(s, t) = (-)$. In words, the only way in which the final health state differs from the initial state is if a $(+)$ matches with a $(-)$, in which case we assume that the $(-)$ becomes infected with probability one.

□ **Payoffs.** Let $u_i(s, h)$ denote player i's utility from the outcome (s, h). We assume the following functional form:

$$u_i(s, h) = V_i(s) - H(h_i),$$

where

$$V_i(s) = \begin{cases} 1 & \text{if } i \in s \\ 0 & \text{if } i \notin s \end{cases}$$

and

$$H(h_i) = \begin{cases} H > 1 & \text{if } h_i = (+) \\ 0 & \text{if } h_i = (-). \end{cases}$$

We interpret $V_i(s)$ as player i's benefit from being matched and $H(h_i)$ as the health costs incurred by i when infected. The assumption that $H > 1$ reflects the obvious truth that AIDS is a serious disease, with a disutility that exceeds the benefits of a single match.

□ **Assessments.** When deciding on a_i^0, player i does not know the history that led to the testing stage. The set of possible histories, $\{(+), (-)\}^3$, is player i's initial information set, I_i^0. Similarly, when deciding on a_i^1, player i does not know the exact history that led to the matching stage. For

[6] Our results continue to hold even if we allow players to decide whether or not to reveal a certificate.

[7] Our results continue to hold even if we allow a player to commit not to match with any player.

every pair (a_i^0, N^C), player i's information set in the matching stage includes all histories that precede the matching stage and in which player i's testing decision is a_i^0, and the set of certified players is N^C. The set of all information sets of player i is denoted $\mathcal{I}_i$. For each $I_i \in \mathcal{I}_i$ we let $X(I_i)$ denote the set of available actions in that history: $X(I_i^0) = \{T, NT\}$ and $X(I_i) = \{C, NC\}$ for $I_i \in \mathcal{I}_i \setminus I_i^0$.

A belief system for player i assigns to every information set of player i a probability measure on the set of histories in the information set. A belief system for player i is denoted μ_i, while a profile of belief systems is denoted μ. A behavioral strategy for player i is a collection of independent probability measures $\beta(I_i)_{I_i \in \mathcal{I}_i}$ with component measures $\beta(I_i)$ defined on $X(I_i)$. A behavioral strategy for player i is denoted β_i, while a profile of behavioral strategies is denoted β. A pair (β, μ) is called an assessment.

□ **The solution concept.** We analyze the sequential equilibria[8] of Γ (see Kreps and Wilson, 1982), which satisfy the following property: given the players' equilibrium testing strategies, there is no profile of matching strategies that Pareto dominates the equilibrium matching strategies. The reason for using this particular refinement of sequential equilibrium is to rule out unreasonable equilibria that result from coordination failure.[9] In what follows, whenever we use the term "sequential equilibrium" (SE), we mean a sequential equilibrium satisfying the above property.

We introduce the following two assumptions that are used in some of the propositions that follow:

$$p < \frac{1}{2} \tag{P1}$$

$$1 - pH > 0. \tag{P2}$$

These assumptions are straightforward to interpret: the first bounds the probability of infection, and the second implies that the expected health costs of a match are low enough that a healthy individual would be willing to match with an untested individual.

We are now in a position to verify the conjecture of Philipson and Posner that certification can result in an equilibrium in which a match always takes place and yet no new infections occur.

Proposition 1. With (P1), there exists an SE with assortative matching: all players test and only players of the same health status match.

Proof. See the Appendix.

Proposition 1 confirms the conjecture of Philipson and Posner in a setting in which there is no fear of the outcome. We now turn to the second part of their conjecture, which asserts that this beneficial outcome is not assured if "many people are fearful of tests which may show they are doomed even if the probability of that result is very low."

3. Incorporating anxiety into the model

■ In this section we incorporate fear of learning bad news into our basic model. Following Philipson and Posner, we assume that there is a disutility associated with pessimistic beliefs about one's health status.[10] We define π_i to be player i's belief about his infection status after the matching stage of the game but prior to the resolution of his actual health status. We introduce a convex "anxiety cost" function $A(\pi_i)$ that gets subtracted from the classical health and matching utilities of the prior section.

[8] Note that we cannot use the notion of perfect Bayesian equilibrium, since the testing decisions are not observable (see Osborne and Rubinstein,1994).

[9] In particular an equilibrium in which two noncertified players would like to match with one another, yet remain unmatched because each has chosen C in the matching stage.

[10] For a discussion of the general difficulties in representing choices of information sources with expected utility over beliefs, see Eliaz and Spiegler (2003).

Our model of individual preferences enables us to capture the anticipatory feelings stressed by Caplin and (2001) while remaining within the revealed-preference framework that dominates current choice theory. The revealed-preference foundation for our model is the following: individuals with our utility function reveal a preference for late resolution of uncertainty by choosing at all stages to avoid any information that is on offer. In technical terms, our utility function is no more than a special case of the model of Kreps and Porteus (1978, 1979) of preferences over the date of resolution of uncertainty. It is the assumed convexity of this anxiety cost function that gives rise to the preference for late resolution of uncertainty.

Although our utility function has classical foundations, these foundations do not extend to our interpretation of $A(\pi_i)$ as anxiety costs. The very same model of individual choice can be motivated using many entirely different psychological interpretations. For example, the preference for late resolution of uncertainty could reflect an innate love of surprise rather than fear of possible bad news. An individual may wish to remain ignorant because it is profoundly pleasurable to live with that tingling feeling of suspense, leaving open the maximum degree of surprise when the uncertainty is finally resolved. The difference between the "fear of bad news" interpretation and the "love of surprise" interpretation has bite only in the discussion of credibility. With the first interpretation, it is clearly difficult for an empathic planner to withhold good news about the outcome of a test. With the second, this news is optimally withheld.

The fact that our decision makers have a preference for late resolution of uncertainty implies that our model cannot be analyzed with classical game-theoretic tools. The payoff depends not only on outcomes, but also on beliefs during the play of the game. In turn, these beliefs depend not only on the prior, but also on the strategy adopted by other players. In particular, the probability of infection when matching with an uncertified player depends on whether or not that player is believed to have tested. Given this, we turn to the theory of psychological games of Geanakoplos, Pearce, and Stacchetti (1989). A psychological game differs from a standard game in that each player's utility depends not only on the game's outcome but also on the player's *beliefs* about the other players. Thus, to turn Γ into a psychological game Γ^P, we need to assign a vector of utility numbers to every terminal history of Γ and every profile of players' beliefs about β.

Let b_i denote player i's beliefs about β_{-i}, and let b denote a profile of such beliefs, one for each player. For every terminal history z, let $a_i^0(z)$ and $a_i^1(z)$ denote player i's testing and matching decisions along z. Given the prior p, and for every terminal history z and profile of beliefs b, define $\pi_i(p, a_i^0(z), a_i^1(z), b_i)$ to be the probability that $h_i = (+)$, given p, z, and b_i (note that in contrast to z and b_i, the prior p is a component of Γ). For notational convenience we write this probability simply as $\pi_i(z, b_i, p)$. Hence, player i's utility from (z, b_i) in a game with a prior of p is defined as follows:

$$u_i(z, b_i, p) = V_i[s(z)] - H[h_i(z)] - A[\pi_i(z, b_i, p)].$$

For ease of notation, we henceforth suppress the argument p in the notation of a player's payoff.[11]

Note that given a profile of beliefs b, the psychological game Γ^P reduces to a standard extensive-form game with imperfect information where each terminal history is associated with a vector of utility numbers. The game associated with a given profile of beliefs b is denoted $\Gamma(b)$. The solution concept we use to analyze Γ^P is sequential psychological equilibrium (SPE). An SPE is a triple (b, β, μ) that satisfies the following properties: (1) (β, μ) is an SE of $\Gamma(b)$, and (2) $b = \beta$.

In what follows, we assume that the cost function $A(\pi_i)$ has the following functional form:

$$A(\pi_i) = \begin{cases} K\left[\dfrac{\pi_i - p}{1 - p}\right]^r & \text{if } \pi_i \geq p, \\ 0 & \text{if } \pi_i < p. \end{cases}$$

[11] Note that unlike Geanakoplos, Pearce, and Stacchetti (1989), in our model the payoff associated with a terminal history and a belief about the other players depends on the prior. The reason for this is that we assume a different domain of preferences than Geanakoplos, Pearce, and Stacchetti. They assume that players have standard expected utility preferences over the set of lotteries on (z, b_i). We, on the other hand, assume that each player has a preference relation over the set of temporal lotteries (as defined by Kreps and Porteus, 1979) on (z, b_i).

In this formulation, the parameter K measures the anxiety that a player experiences when he learns that he is infected for sure. The parameter r captures the agent's fear of learning "really bad news" relative to learning "bad news." To see this, note that the utility loss from any partial resolution of uncertainty (e.g., an imperfect test) goes to zero as r goes to infinity.

With this modification of the game, we are in a position to demonstrate that assortative matching may fail when players exhibit a preference for late resolution of uncertainty. In particular, we identify sufficient conditions for ensuring a unique equilibrium in which no player tests and matching occurs with certainty. Thus, under these conditions, infections *must* arise with positive probability.

Proposition 2. Assume (P1) and (P2) hold. If $pK > 1$ and $p^r K \leq 1 - p(1-p)H$, then there exists a unique SPE in which no player tests and matching occurs with certainty.

Proof. See the Appendix.

Note that the first additional condition in the proposition, $pK > 1$, ensures that the expected gain from matching can never be enough to compensate for the anxiety that one expects to incur from learning that one is infected for sure. The second additional condition ensures that r is high enough that the anxiety caused by matching with another untested individual is not overwhelming. Together these conditions place us in the world envisaged by Philipson and Posner: untested individuals are willing to match with one another, and the anxiety caused by learning that one is infected is unbearable. It is hardly surprising then that society finds itself in a steady state with continuing infections.

It should be noted that the parametric assumptions we make are not delicate, and are not designed for realism. In particular, the assumption that fear overwhelms matching benefits is very strong. If the social equilibrium involved all others testing, surely the holdouts would be forced into testing. What this suggests is that an implementation-theoretic approach may ultimately be of greater value than the mechanism-design approach of this article. The struggle may be to move society away from a bad equilibrium and toward a superior equilibrium that is currently not being played. Yet, as we shall point out in Section 6, adoption of this approach comes at a high cost in terms of complication. Our goal in the current article is to make as clear as possible our fundamental point concerning the role of psychological incentives in the spread of AIDS, and it is for this reason that we are willing to make strong assumptions on the underlying parameters.

4. Policy: the framework

■ **Messages and mechanisms.** In the game described above, the certification technology is perfectly revealing of the true state of the world. In this section we define a class of message-transmission technologies that are not perfectly correlated with the true state. We refer to this broader class of message-transmission technologies as *mechanisms*. In principle, such mechanisms may allow for any number of distinct messages to be provided to those who test and also allow the planner to randomize over the entire set of message vectors.

For the bulk of our analysis it is adequate to restrict attention to *simple* mechanisms that may mix over only two types of messages: a certificate of some sort, C, and a null message, $\emptyset$ (to the outside observer, one who receives this message will be indistinguishable from one who did not test). The psychological game induced by a mechanism g is denoted $\Gamma^P(g)$.

There are three important assumptions we make with regard to mechanisms:

(i) Certificates cannot be forged: If a player reveals a certificate, that certificate could only have come from the mechanism.

(ii) The planner is able to commit to a mechanism.

(iii) The mechanism is common knowledge.

Although our "message-transmission" approach is borrowed from mechanism-design theory, our policy problem is quite different. One difference is that in mechanism design, the policy maker

is at an informational disadvantage. Private agents know their own type, the policy maker does not. In our model, the private agents have no private information.

A second difference is that in mechanism design, the policy maker is free to design the game form in which the agents interact. The only constraint the planner faces is that the mechanism she designs must meet a "participation constraint." The typical participation constraint takes the following form: in the desirable equilibrium, each agent's payoff should be at least as high as his *exogenous* reservation utility. In our model the planner does not design the game form in which the agents interact: this game form is exogenously given. The only way the planner can influence the agents' behavior is by affecting their beliefs through the messages that she sends. As in mechanism design, the planner cannot force the agents to participate in her mechanism, in the sense that she cannot force her messages on the agents. The agents must find it optimal to obtain information from the planner. However, in contrast to mechanism design, if an agent refuses to participate in the mechanism (i.e., if an agent refuses to test), then his payoff is determined *endogenously* through the interaction with the other agents.

□ **Conditional and unconditional mechanisms.** We focus on two different classes of mechanism: those that can condition the procedure of sending messages on the test results of all individuals, and those that have less flexibility in their choice of messages. We refer to the broader class as the set of conditional mechanisms and to the narrower as unconditional mechanisms. There is one global assumption we make to capture the idea that no message can be given to an individual who does not show up for the test. The assumption also ensures that a player with an unattractive message can pretend not to have tested at all (however, in equilibrium players may infer that this is likely to be a pretense and update beliefs accordingly). This feasibility constraint is all that we impose on the class of conditional mechanisms.

No shows: A player who does not test is certain to receive the null message. (M1)

A conditional mechanism can respond differently to an infected individual depending on the test results of surrounding players. As we show below, this additional flexibility may have value. In practice, of course, it may be infeasible to display this form of flexibility. The conditional mechanism requires all messages to be sent only once all tests have been completed, and this may be impractical if tests are not taken simultaneously. There may also be exogenous constraints (legal, moral, or political) that prevent the planner from discriminating between two individuals with identical test results. We therefore introduce unconditional mechanisms that satisfy additional equal treatment and independence requirements.

Equal treatment of test results: If $t_i = t_j$, then for all $m \in \{C, \emptyset\}$,

$$\Pr(m_i = m) = \Pr(m_j = m). \tag{M2}$$

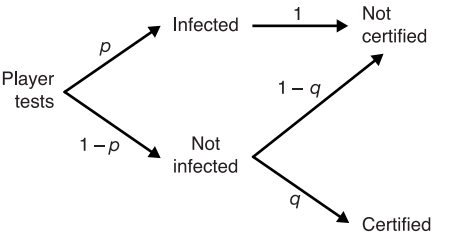
Independence: If player i tests, then the probability that $m_i = m$ depends only on player i's type and actions. (M3)

5. Implementing zero infections

■ The question on which we focus is how to design policies that have as their dominant goal the prevention of infection. In the language of implementation theory, we assume that the social choice rule of the planner associates with every vector of types the set of outcomes in which no infections occur. Hence our goal is to identify circumstances in which the policy maker can design a mechanism g that yields an equilibrium with no infections even when (P1) and (P2) hold. A mechanism that achieves this is said to (weakly) implement zero infections.

□ **Unconditional mechanisms.** Assume (P1) and (P2) hold. This means that without intervention, infections occur with positive probability. But if the disutility from being infected is sufficiently high, we identify in Proposition 3 a simple mechanism that implements zero infections. This mechanism sends a certificate of health to some proportion of those who are in fact healthy, but never sends a certificate to those who are infected. We refer to unconditional mechanisms of

FIGURE 1



this type as *certification schemes*, since they are so strongly analogous to the certification policies discussed by Philipson and Posner.

Proposition 3. Assume (P1), (P2), and $H > 4$. Then there exists $\overline{r} > 1$ such that for all $r \geq \overline{r}$, the no-infections outcome is weakly implementable by a certification scheme.

Proof. See the Appendix.

The intuition for the proof is as follows. What discourages a player from testing is the fear of learning that he is infected with certainty. However, testing may also result in good news, which provides a competitive advantage in the matching game. So to encourage a player to test we need to ensure that on the one hand, a tested player never receives unambiguously bad news, and on the other hand, there must be a message that conveys unambiguously good news (to ensure zero infections). This can be achieved by a mechanism that certifies only individuals who test negative but does not certify all such individuals. In other words, a player who tested negative is certified with some positive probability q that is strictly less than one, while a player who tested positive is certain to remain uncertified. Figure 1 illustrates the way the proposed mechanism works.

In the proof, we set the probability that a noninfected individual is certified, q, equal to $[1/(1 - p)](1 - pH/2)$. This implies that the probability that a noncertified player is infected is equal to $2/H$. If r is sufficiently large, then partial resolution of uncertainty leads to only a small loss of utility. Hence, the increased "anxiety" that a player experiences when he decides to test is offset by the increase in matching and health benefits. This ensures the existence of an equilibrium in which all players test. By our choice of q, any player prefers to remain unmatched rather than to match with a player who tested but was not certified. This implies that only certified players, who are healthy for sure, will end up matching. Hence, no infections arise.

The certification scheme of Proposition 3 operates by leaving uncertain some who are in fact free of infection. One can imagine an actual medical test that for scientific or technological reasons has exactly this property: there exists a grey area in which the test cannot provide either a positive or negative result with certainty. In that context, our result can be read as carrying an implicit suggestion for medical researchers to develop tests that are particularly good at avoiding false negatives. Note that Proposition 3 can also be interpreted as saying that if individuals prefer to delay the resolution of uncertainty, then one way to battle infections is to ration health certificates. As explained above, the particular method used to ration the certificates is important in order to ensure that people test and that only healthy individuals match.

□ **Conditional mechanisms.** It is clear from the proof of Proposition 3 that there is a connected set of certification schemes with varying levels of *ex post* beliefs that satisfy all of the required conditions for there to be no new infections. Of course, there are limits to what can be achieved with these mechanisms. In particular, the planner will be unable to guarantee matches 100% of the time, since this would require her to use a fully revealing test, in which case Proposition 2 has already established that under (P1) and (P2) every equilibrium involves infections. In this section we show that the feasible set of policies is substantially richer when we allow for conditioning. To demonstrate this, we focus on a specific mechanism, which we call the Minimally Informative Guidance mechanism (MIG). Under certain additional conditions, the planner will be able to use

this mechanism not only to keep infections to zero, but also to generate full trade. Furthermore, she will be able to keep anxiety at its absolute minimum consistent with these two outcomes.

Definition 1. The MIG mechanism is a conditional mechanism with the following rules:

> If there are two individuals with identical test results, then only those individuals are certified. If there are three individuals with identical test results, then only two, who are randomly chosen, are certified. (R1)

> In all other cases, an individual is certified if and only if he is healthy. (R2)

Proposition 4. Assume that (P1) and (P2) hold. If $H > 4$ and $(2p)^r(1-2p)^r \leq 2/K$, then the MIG mechanism weakly implements an outcome with the following properties:

(i) a pair of players with the same health status always match, and

(ii) conditional on (i), the sum of expected utilities is maximized.

Proof. See the Appendix.

The intuition for the proof is as follows. Given that all players test, (R1) guarantees two things: (1) a match will always take place, and (2) players who match have the same health status. Given (1) and (2), an unmatched player must infer that most likely his health status is different from that of the matched players. Since $p < 1/2$, it is more likely that he, the unmatched player, is infected while the other two are not. Clearly, such "bad news" is necessarily transmitted by any mechanism that implements zero infections. Hence, conditional on having no infections and sure matches, the MIG mechanism maximizes the players' *ex ante* expected utility.

The role of (R2) is to ensure that deviations from testing are not profitable. With our assumptions, one who deviates to not testing ends up certain to remain single. Against this, the only cost of testing is the possibility of not being certified. However, the MIG mechanism is constructed in such a way that the probability of not being certified is relatively small, and the information conveyed by the absence of certification is not that bad.

6. Extensions

■ We briefly outline three important extensions of our model.

□ **Credibility.** Our mechanisms suffer from credibility problems if we envisage test results as delivered by a human policy maker rather than by a machine. To formally analyze credibility, we introduce the planner as a player in the policy game that she designs, as in Baliga, Corchon, and Sjöström (1997). Suppose we modify the psychological game Γ^P such that the planner, rather than nature, moves after the agents have made their testing decisions. The planner's strategy is defined to be a choice of a message-transmission technology that satisfies (M1). We assume that the planner has lexicographic preferences: her primary concern is to have zero infections, her secondary concern is to maximize the sum of agents' utilities.[12]

Consider a planner who, upon observing the individuals' test results, needs to decide which messages to send. Imagine all three individuals are found to be healthy. It seems reasonable to expect the planner to certify them all. This clearly violates the rules of both the MIG and the unconditional mechanisms. Hence, intuition suggests that both types of mechanisms discussed in this article are not credible.

To formally express the above intuition, recall that a mechanism g weakly implements an outcome if that outcome can be attained in some equilibrium of the game. That equilibrium is said to be the desirable equilibrium associated with the mechanism g and is denoted $e(g)$. Given this notation, it is natural to call a mechanism g credible if the extended game, in which the planner is

[12] Note that given the agents' beliefs, the planner can choose to "lie" to the agents (e.g., given that agents believe that only healthy individuals are certified, the planner can choose to certify all agents). Hence we assume that the planner evaluates the agents' utilities based on their beliefs, which may be incorrect. For a discussion of the theoretical implications of this issue, see Caplin and Leahy (forthcoming).

a player, has an equilibrium with the following properties: (i) the agents' strategies are the same as in $e(g)$, and (ii) the planner's strategy coincides with the rules of g.

It turns out that this definition is not completely satisfactory for our model because of a subtlety involving out-of-equilibrium beliefs. To see the issue, suppose all three individuals believe that the planner behaves according to the MIG mechanism. In addition, suppose that individuals believe that whenever they are all certified, then they all must be infected for sure. Given these beliefs, a planner would not want to certify all individuals even if all are found to be healthy. With this example in mind, we add an additional requirement to the definition of credibility: that there exists no out-of-equilibrium belief for which the planner would want to deviate from the rules of the mechanism.

Observation 1. The MIG mechanism and the certification scheme of Proposition 3 are both *not* credible.

Proof. The proof follows the intuitive argument given above. Assume the MIG is credible. Then there exists an SPE in which all individuals test and the planner's strategy satisfies (R1) and (R2) of Definition 1. Moreover, there are no out-of-equilibrium beliefs for which the planner would want to violate (R1) or (R2). Consider the history in which all individuals are healthy, they all test, and the planner certifies them all. Because the planner's action in this history violates (R1), this history clearly describes an out-of-equilibrium event. Consider the players' information sets in the matching stage that follows the above history. Suppose that at these information sets, each individual assigns probability one to the history in which all individuals are healthy and all decided to test. Following exactly this history and given the above beliefs, the optimal action for the planner is to certify all individuals. This violates (R1) and contradicts our assumption that the MIG is credible.

Assume next that the certification scheme of Proposition 3 is credible. Recall that this mechanism has the property that whenever a healthy individual tests, he receives the null message with strictly positive probability. Using the same arguments given above for the MIG mechanism, it follows that for some out-of-equilibrium beliefs, the planner would want to violate the rules of the certification scheme whenever all individuals are healthy and test. *Q.E.D.*

One resolution to the problem of credibility is to mechanize the message-transmission technology. In this respect it seems that the technology of the certification scheme is easier to mechanize than that of the MIG. Recall that unlike the MIG, the unconditional certification scheme can hand out certificates one at a time, as individuals test. In contrast, the MIG requires waiting until all tests have been taken before deciding which certificates to hand out, and to whom.

The MIG mechanism has another key weakness that differentiates it from the unconditional mechanism. The MIG works only as long as each person's health status is never revealed. If a person's health status becomes known for some reason, then all the other individuals could infer their true health status. If the players anticipate this, they might not want to test.

□ **Large populations.** Although we focus on a market with three players, we conjecture that our results continue to hold in a case with many players. To extend our analysis to accommodate a large population of agents, we need to modify both the testing and matching stages as follows. After players privately decide whether or not to test, they are informed of the *fraction* of players who were certified. Each player then decides on his matching strategy. Pairs of players are then randomly matched, and only eligible pairs exit the market. Unmatched players continue to search through pairwise matching until they are matched, or until they exit the market unmatched (either by choice or after meeting all remaining players). By assuming that search is costly, one can obtain a competitive advantage to being certified similar to the one we have in our model. The analysis becomes much more involved as one needs to keep track of how players update their beliefs during the matching process.

□ **Full implementation.** In this article we used the mechanism-design approach, thereby focusing on the problem of generating at least one equilibrium with no infections. One possible

criticism of this approach is that there may very well be other undesirable equilibria in which infections occur. This problem is dealt with by implementation theory, which studies the design of mechanisms that generate unique desirable equilibria (see Jackson (2001) for a recent survey of this literature).

Full implementation of the no-infections outcome is the subject of our working paper, Caplin and Eliaz (2002). The cost in terms of complexity of pursuing this approach turns out to be very high. The required game form has many more stages than the one we present here. In particular, agents choose whether or not to reveal certificates, and the matching stage is modelled as a sequential alternating offer game. We show that under certain conditions, the certification scheme of Proposition 3 fully implements the no-infections outcome. We also show that under certain conditions, a modified MIG mechanism generates a unique equilibrium outcome with the properties described in Proposition 4. For detailed proofs of these results, please consult our working paper.

7. Concluding remarks

■ We have built a simple model allowing us to confirm that psychological interventions may help slow the spread of AIDS. While the model is rudimentary, its potential for further development seems clear.

Appendix

■ Proofs of Propositions 1–4 follow.

Proof of Proposition 1. Let β^* be the following profile of (pure) symmetric behavioral strategies: each player tests, certified players choose C, and noncertified players choose NC. If $H > 1/(1-p)$, then an untested player chooses C; otherwise, he chooses NC.

Let μ^* be a belief system with the following properties. Prior to the testing stage, each player believes that the probability that any one player is infected is p. Following the testing stage, any player who tests has his beliefs updated by the outcome of this test. Any player who does not test has no change in beliefs. At the matching stage, each player believes that the other players have tested. Hence, any player without a certificate is believed to be type (+) with certainty.

It is immediate that the behavior dictated by β^* in the matching subgame produces the assortative matching outcome. We now verify the sequential rationality of β^*. We begin by verifying the sequential rationality of the behavioral strategies in the matching stage. First, it is sequentially rational for a player who tested positive to choose NC. Any match produces the matching benefit at no cost in terms of health. Second, it is also sequentially rational for a certified individual to choose C, since by μ^*, a certified individual is healthy whereas a noncertified individual is infected. It remains to verify the sequential rationality of the behavioral strategy of an untested player. If $H > 1/(1-p)$, then by μ^*, a match with a noncertified individual is too costly in terms of health even for an untested individual. If $H \leq 1/(1-p)$, then the benefits from matching exceed even the expected cost of getting infected.

We now verify that testing at the start of the game is sequentially rational. Assume first that $H > 1/(1-p)$, and suppose player i decides to deviate from β^* by not testing. Given β^*_{-i}, player i would be able to match only with an infected individual. From $H > 1/(1-p)$ it follows that the highest payoff an untested player can expect to receive is $-pH$. Suppose player i adheres to his equilibrium strategy. Then, first, he would never be infected. Hence, his expected health costs are $-pH$. Second, player i will match with probability $2/3$. It follows that testing is sequentially rational.

Assume next that $H \leq 1/(1-p)$, and suppose player i decides to deviate from β^* by not testing. Player i can match only with infected players. From $H \leq 1/(1-p)$ it follows that it is optimal for him to do so. Hence, the highest expected payoff that i can get by not testing is

$$\left[\frac{2}{3}p^2 + 2p(1-p)\right] - pH - (1-p)H\left[\frac{2}{3}p^2 + 2p(1-p)\right]. \tag{A1}$$

By testing, player i can obtain an expected payoff of

$$\frac{2}{3} - pH. \tag{A2}$$

For $p < 1/2$ we have that expression (A2) is greater than expression (A1). Hence, it is not profitable to deviate from testing.

To verify the consistency of (β^*, μ^*), let the completely mixed strategy β^ε be a perturbation of β^* such that, for each

player i and at each information set, feasible actions given probability zero in β^* are given (small) probability $\varepsilon > 0$, while the action selected according to β^* is given the residual probability. Now define μ^ε to be the (well-defined) belief system derived from β^ε via Bayes' rule. Clearly, $(\beta^\varepsilon, \mu^\varepsilon)$ converge to (β^*, μ^*) as $\varepsilon \to 0$, confirming consistency. *Q.E.D.*

Proof of Proposition 2. Assume (P1) and (P2) and that the following inequalities hold:

$$pK > 1 \tag{A3}$$

and

$$p^r K < 1 - p(1-p)H. \tag{A4}$$

By (A3), any strategy that assigns a positive probability to testing is strictly dominated. By (A4), an untested player strictly prefers to match rather than stay unmatched. Hence, an untested player strictly prefers NC to C. By appropriately constructing the players' beliefs, it is straightforward to show that there exists an SPE in which no player tests and matching occurs with certainty. Note that our refinement of SPE rules out an equilibrium in which no player tests but every player chooses C in the matching stage. *Q.E.D.*

Proof of Proposition 3. Let g be a certification scheme with the property that for every i,

$$\Pr(h_i = (+) \mid m_i = \emptyset) = \frac{2}{H} < \frac{1}{2},$$

where the inequality follows from our assumption that $H > 4$ (note that $p < 2/H$ by (P2), hence this mechanism is feasible).

Let (b^*, β^*, μ^*) be the following candidate for an SPE in the game $\Gamma^P(g)$. The strategy profile β^* satisfies that each player tests and chooses C at any information set in the matching stage. We set $b^* = \beta^*$.

To describe the belief system μ^* we introduce the following notation. Let $P_i[t_j = (+)]$ be the (unconditional) probability that player i assigns to the event that player j is of type (+). At the initial information, $P_i[t_j = (+)] = p$ for every i. Consider next an information set following a history (t, a^0, N^C). The beliefs of player i about his own type are given by

$$\begin{aligned} P_i[t_i = (+) \mid a_i^0 = NT,\ i \notin N^C] &= p, \\ P_i[t_i = (+) \mid a_i^0 = T,\ i \notin N^C] &= \frac{2}{H}, \\ P_i[t_i = (+) \mid a_i^0 = T,\ i \in N^C] &= 0. \end{aligned}$$

For any $a_i^0 \in \{T, NT\}$, the beliefs of player i about the type of player $j \neq i$ are given by

$$\begin{aligned} P_i[t_j = (+) \mid a_i^0,\ j \notin N^C] &= \frac{2}{H}, \\ P_i[t_j = (+) \mid a_i^0,\ j \in N^C] &= 0. \end{aligned}$$

It is immediate that the behavior dictated by β^* in the matching subgame produces zero infections. It remains to show that there exists $\bar{r} \geq 1$ such that for all $r \geq \bar{r}$, (β^*, μ^*) is an SE of $\Gamma(b^*)$. Consistency follows precisely as in Proposition 1. We therefore proceed to the proof of sequential rationality.

Claim A1. With (P1) and $H > 4$, certified/noncertified/untested individuals want to match only with certified individuals.

Proof. Player i prefers to match with a noncertified player than to remain single if and only if

$$1 \geq (1 - x_i)\left(\frac{2}{H}\right) H = 2(1 - x_i), \tag{A5}$$

where

$$x_i = P_i\left[t_i = (+) \mid a_i^0, N^C\right].$$

For a certified player, $x_i = 0$. For a noncertified player, $x_i = 2/H$, which is smaller than $1/2$ by our assumption that $H > 4$. Finally, for an untested player, $x_i = p$, which by assumption (P1) is smaller than $1/2$. It follows that (A5) does not hold. *Q.E.D.*

Claim A2. There exists $\bar{r} \geq 1$ such that for $r \geq \bar{r}$, no player can gain by deviating to not testing.

Proof. According to the profile of strategies β^*, whether or not a player tests has no impact on health costs. However, a player who does not test never matches. Therefore, given that the other players test, player i prefers to test if the probability that he matches is at least as high as the expected "anxiety" from testing.

Player i matches only if he is certified and at least one other player is also certified. If he tests, his probability of being certified is $1 - pH/2$. Conditional on being certified, he matches for sure if exactly one other player is also certified. If the two other players are both certified, then he matches with probability $2/3$. Hence, the probability that he matches in equilibrium is given by the following expression:

$$2\frac{pH}{2}\left(1-\frac{pH}{2}\right)^2+\frac{2}{3}\left(1-\frac{pH}{2}\right)^3.$$

This can be rewritten as follows:

$$\frac{2}{3}(1+pH)\left(1-\frac{pH}{2}\right)^2. \tag{A6}$$

Player i incurs a cost of anxiety only when he is not certified. This event occurs with probability $pH/2$. By the construction of β^*, if player i is not certified, he remains single. Hence, conditional on not being certified, the probability that player i will end up infected is equal to $2/H$. It follows that the expected cost of anxiety is given by the expression

$$\frac{pH}{2}K\left(\frac{\frac{2}{H}-p}{1-p}\right)^r. \tag{A7}$$

To define $\bar{r}$, first evaluate expression (A7) at $r = 1$. If this is less than or equal to expression (A6), set $\bar{r} = 1$. If not, set $\bar{r}$ such that

$$\frac{pH}{2}K\left(\frac{\frac{2}{H}-p}{1-p}\right)^{\bar{r}}=\frac{2}{3}(1+pH)\left(1-\frac{pH}{2}\right)^2.$$

By construction, for every $r \geq \bar{r}$, expression (A6) is greater than expression (A7), establishing the claim. *Q.E.D.*

Taken together, Claims A1 and A2 verify that for all $r \geq \bar{r}$, β^* is sequentially rational. This completes the proof of Proposition 3. *Q.E.D.*

Proof of Proposition 4. We proceed in two steps. First, we verify the existence of an SPE in which no infections occur and matching always takes place. Then we show that the expected anxiety of each player is the minimum possible, subject to having no infections and sure matches. This ensures that the sum of the agents' expected utilities is maximized subject to having assortative matching in equilibrium.

Step 1: existence. Let (b^*, β^*, μ^*) be the following candidate for an SPE. Let β^* be a profile of behavioral strategies such that for every player i,

(i) $a_i^0 = T$,
(ii) $a_i^1 = NC$ if $a_i^0 = T, i \notin N^C$, and $|N^C| \leq 1$, and
(iii) $a_i^1 = C$ in all other cases.

We set the profile of beliefs b^* to be equal to β^*.

The belief system μ^* is constructed as follows. Consider player i and the information sets in which $a_i^0 = T$"

(a) If $|N^C| = 2$, then for every $j \in N^C$, the probability that $t_j = (+)$ is given by the expression

$$\frac{2p^3+3p^2(1-p)}{2p^3+3p^2(1-p)+3p(1-p)^2+2(1-p)^3}. \tag{A8}$$

Note that with (P1), (A8) is smaller than p. The probability that $t_j = (+)$ for $j \notin N^C$ is given by

$$p^3+3p(1-p)^2. \tag{A9}$$

Note that (A9) is larger than p.

(b) If $N^C = \{j\}$, where $j \neq i$, then player i believes that he is infected, player j is healthy, and player k is untested and therefore infected with probability p.

(c) If $N^C = \{i\}$, then player i believes that exactly one other player has not tested. He therefore believes that he is healthy for sure and that each of the other players is infected with probability $(1/2)(1 + p)$.

(d) If $N^C = \emptyset$, then player i believes that he is infected for sure and both other players have not tested. He therefore believes that each of them is infected with probability p.

Consider next the information sets in which $a_i^0 = NT$. After every history, which is consistent with the mechanism, player i believes that he is infected with probability p. The probability that player i assigns to the event that $t_j = (+)$ for some $j \neq i$ is conditioned on N^C. If $N^C = \{j, k\}$, then player i believes that both players tested and received identical results. Hence, $t_j = (+)$ with probability $p^2/[p^2 + (1-p)^2]$. If $N^C = \{j\}$, then player i believes that j is healthy for sure, while k is infected for sure. If N^C is empty, then player i believes that exactly one player has not tested, while another player tested and is infected for sure. He therefore believes that each of the other players is infected with probability $(1/2)(1 + p)$.

Clearly, if players behave according to β^*, then with probability one a pair of players (i, j) with $t_i = t_j$ would match. It remains to show that our assumptions imply that (β^*, μ^*) is an SE of $\Gamma(b^*)$. Consistency follows precisely as in Proposition 1. We therefore proceed to the proof of sequential rationality.

We begin by verifying that the matching strategies are sequentially rational for all $r > 1$. Consider first the information sets with histories in which $a_i^0 = T, i \notin N^C$, and $|N^C| \leq 1$. In each of these information sets $t_i = (+)$. Hence, $a_i^1 = NC$ is weakly dominating. Next consider all other information sets. We claim that in each of these information sets it is optimal for player i to choose C. To see why, consider the different realizations of N^C:

(a) $N^C = \{i, j\}$. According to μ^*, player j has a higher chance of being healthy than the noncertified player k. According to $\beta_j^*, a_j^1 = C$. Hence, it is optimal for player i to set $a_i^1 = C$.

(b) $N^C = \{j, k\}$. According to $\beta_j^*, a_j^1 = a_k^1 = C$. Hence, with probability one, player i will not match. Therefore, he is indifferent between C and NC.

(c) $N^C = \{i\}$. According to μ^*, player i believes that he is healthy for sure while each of the other players is infected with probability $(1/2)(1 + p)$. By remaining unmatched, player i expects to receive a payoff of zero. By matching with one of the other players he expects to receive a payoff of

$$1 - \frac{1}{2}(1 + p)H - \left(\frac{1}{2}\right)^r K. \tag{A10}$$

With $H > 4$, it follows that (A10) is negative. Hence, it is optimal for player i to remain unmatched. To do this, player i must choose C in the matching stage.

(d) $N^C = \emptyset$. We have already considered the case of $a_i^0 = T$. Assume then that $a_i^0 = NT$. According to μ^*, player i believes that each of the other players is infected with probability $(1/2)(1 + p)$. By remaining unmatched, player i expects to receive a payoff of $-pH$. By matching with one of the other players, he expects to receive a payoff of

$$1 - \left[p + (1-p)\frac{1}{2}(1 + p)\right] H - \left(\frac{1}{2}\right)^r K.$$

Therefore, player i is better off not matching if and only if this expression is negative. From our argument for the case in which $N^C = \{i\}$, it follows that it is optimal for player i to choose C.

We now show that the testing decisions are sequentially rational. If player i follows his equilibrium strategy, then he matches with probability $2/3$, he incurs an expected health cost of pH, and his expected cost of anxiety is $(1/3)K(2p)^r(1-2p)^r$. Suppose instead that player i deviates to NT. Then it is optimal for him to choose C. To see this, note first that according to β_{-i}^*, any player $j \neq i$ who certifies will choose $a_j^1 = C$. Second, any player who is not certified must be infected for sure. It follows that the highest payoff that player i can obtain by not testing is $-pH$. Thus, deviation is not profitable if and only if

$$\frac{2}{3} - \frac{1}{3}K(2p)^r(1-2p)^r \geq 0. \tag{A11}$$

Inequality (A11) is satisfied by assumption, which implies that testing is sequentially rational. This completes the proof of Step 1.

Step 2: minimal anxiety. A necessary condition to have a match in every contingency and to eliminate the risk of infections is the following: in every contingency, two players of the same health status are matched. Since $p < 1/2$, any unmatched player necessarily infers that he is more likely to be infected than the players who are matched. The expected probability that an unmatched player is infected is given by (A9). Similarly, a matched player necessarily infers that being matched does not guarantee that one is healthy. Thus, the expected probability that a matched player is infected is given by (A8). The only thing that a more elaborate mechanism could do would be to provide additional information. If it provides more information, however, all it can do is perform mean-preserving spreads around these beliefs, which would necessarily raise anxiety in light of Jensen's inequality.

This completes the proof of Proposition 4. *Q.E.D.*

References

BALIGA, S., CORCHON, L.C., AND SJÖSTRÖM, T. "The Theory of Implementation When the Planner Is a Player." *Journal of Economic Theory*, Vol. 77 (1997), pp. 15–33.

CAPLIN, A. "Fear as a Policy Instrument: Economic and Psychological Perspectives on Intertemporal Choice." In G. Loewenstein, D. Read, and R. Baumeister, eds., *Time and Decision*. New York: Russell Sage Foundation, 2003.

——— AND ELIAZ, K. "AIDS Policy and Psychology: An Implementation Theoretic Approach." Mimeo, Department of Economics, New York University, 2002.

——— AND LEAHY, J. "The Supply of Information by a Concerned Expert." *Economic Journal*, forthcoming.

——— AND ———. "Psychological Expected Utility Theory and Anticipatory Feelings." *Quarterly Journal of Economics*, Vol. 116 (2001), pp. 55–79.

——— AND ———. "Behavioral Policy." In I. Brocas and J.D. Carrillo, eds., *The Psychology of Economic Decisions: Vol. 1*. New York: Oxford University Press, 2003.

COHEN, J. "Tough Guys: Men Avoid Preventive Health Care in Sickness and in Health." http://abcnews.go.com/sections/living/DailyNews/doctordelay060702.html, 2002.

ELIAZ, K. AND SPIEGLER, R. "Anticipatory Feelings and Choices of Information Sources." Mimeo, New York University and Tel-Aviv University, 2003.

GEANAKOPLOS, J., PEARCE, D., AND STACCHETTI, E. "Psychological Games and Sequential Rationality." *Games and Economic Behavior*, Vol. 1 (1989), pp. 60–79.

JACKSON, M.O. "A Crash Course in Implementation Theory." *Social Choice and Welfare*, Vol. 18 (2001), pp. 655–708.

KREMER, M. "Integrating Behavioral Choice into Epidemiological Models of AIDS." *Quarterly Journal of Economics*, Vol. 111 (1996), pp. 549–573.

KREPS, D.M. AND PORTEUS, E.L. "Temporal Resolution of Uncertainty and Dynamic Choice Theory." *Econometrica*, Vol. 46 (1978), pp. 185–200.

——— AND ———. "Temporal von Neumann-Morgenstern and Induced Preferences." *Journal of Economic Theory*, Vol. 20 (1979), pp. 81–109.

——— AND WILSON, R. "Sequential Equilibria." *Econometrica*, Vol. 50 (1982), pp. 863–894.

LYTER, D., VALDISERRI, R., KINGSLEY, L., AMOROSO, W., AND RINALDO, C. "The HIV Antibody Test: Why Gay and Bisexual Men Want or Do Not Want To Know Their Results." *Public Health Reports*, Vol. 102 (1987), pp. 468–474.

OSBORNE, M.J. AND RUBINSTEIN, A. *A Course in Game Theory*. Cambridge, Mass.: MIT Press, 1994.

PHILIPSON, T. "Economic Epidemiology and Infectious Diseases." In A.J. Culyer and J.P Newhouse, eds., *Handbook of Health Economics*. Amsterdam: North-Holland, 2000.

——— AND POSNER, R.A. "A Theoretical and Empirical Investigation of the Effects of Public Health Subsidies for STD Testing." *Quarterly Journal of Economics*, Vol. 110 (1995), pp. 445–474.

WITTE, K. "Fear as Motivator, Fear as Inhibitor: Using the Extended Parallel Process Model to Explain Fear Appeal Successes and Failures." In P.A. Andersen and L.K. Guerrero, eds., *The Handbook of Communication and Emotion: Research, Theory, Applications, and Contexts*. San Diego: Academic Press, 1998.

——— AND ALLEN, M. "A Meta-Analysis of Fear Appeals: Implications for Effective Public Health Campaigns." *Health Education and Behavior*, Vol. 27 (2000), pp. 591–615.