Wasonian Persuasion*

Kfir Eliaz

School of Economics, Tel Aviv University

and

Ariel Rubinstein

School of Economics, Tel Aviv University
and Department of Economics, New York University

April 20, 2025

Abstract

A firm wishes to persuade a patient to take a drug by making either positive statements like "if you take our drug, you will be cured", or negative statements like "anyone who was not cured did not take our drug". Patients are neither Bayesian nor strategic: They use a decision procedure based on sampling past cases. We characterize the firm's optimal statement, and analyze competition between firms making either positive statements about themselves or negative statements about their rivals. The model highlights that logically equivalent statements can differ in effectiveness and identifies circumstances favoring negative ads over positive ones.

*We are indebted to Áron Tóbiás who carefully went through the draft of the paper and saved us from many mistakes. Thanks also to Ran Eilat, Michael Richter and Rani Spiegler for their comments.

1 Introduction

A drug manufacturer promises that its drug will cure your disease. A personal trainer warns that if you don't exercise daily your health will deteriorate. A local auto mechanic claims that any neighbour who had an accident did not use his services.

In the above examples, a party is making a statement in order to persuade you to take a particular action. The statement describes a relation between an action and a consequence, i.e., how one implies the other. If you believe the statement is true, you will take the action; otherwise, you won't. How do we assess such statements?

In contrast to what is usually assumed in economic models, we typically don't have a prior on the validity of the statement, and furthermore, the ability to think strategically is limited. A common practice is to sample a small number of relevant past cases, i.e., to collect data on individuals who faced the same dilemma in the past including their decision (e.g., whether they took the drug, exercised, or used the local auto mechanic) and the subsequent outcomes (e.g., whether they were cured, their health, and any involvement of their car in accidents). Based on this limited data, one assesses the validity of the statement and comes to a decision.

The leading example throughout the paper is a firm that offers a drug to cure a certain disease. Two exogenous parameters capture the effectiveness of the drug: the probability p that a patient will be cured if he takes the drug, and the probability q that he will get well without it. Both parameters are known to the firm but not to the patients. The firm can make one of four statements which all have the form $A \rightarrow B :$ two positive ones - "taking the drug will cure you" and "patients who were cured took the drug" - and two negative ones - "not taking the drug will result in no cure" and "people who were not cured, had not taken the drug".

Modeling the process by which patients assess the validity of a statement requires specifying the following:

(i) *On what information does a patient base his assessment?* We focus on two types of data collection procedures which are used by patients to search past cases, where it is known whether a patient took the drug and whether he was cured. One involves sampling a finite number of past cases while another involves sequentially sampling past cases until a relevant case is identified.
(ii) *What data do patients find relevant when assessing a statement of the form $A \rightarrow B$?* We assume that a refutation (a case where A is true and B is false) and a confirmation (a case where both A and B are true) are relevant (in Section 4 another type of case may also be relevant).
(iii) *How are the relevant cases that are sampled used to determine the decision?* We assume that a refutation leads to rejection of the firm's drug while a confirmation without a refutation persuades patients to take the drug.

The question of which information is relevant in assessing the validity of an implication statement is related to the Wason Selection Task experiment (Wason, 1960).[1] The title "Wasonian Persuasion" is a nod to this experiment. In Wason's experiment (and its variations) subjects exhibit a tendency to regard information as relevant if it confirms the statement and not only if it falsifies it (see Wason and Johnson-Laird, 1972, and Klayman and Ha, 1987). This is likely because the everyday meaning of an implication statement "if A then B" differs from its "logical" definition. When we say "anyone who takes the drug will be cured", we typically refer to the population in general (unlike in Wason's experiment) and we mean to say that:
(i) the drug *always* cures;
(ii) there are cases of patients who took the drug and were cured; and
(iii) anyone who does *not* take the drug is significantly *less* likely to be cured.

[1] In this experiment, subjects are asked: "Suppose each card has a number on one side, and a letter on the other. There are four cards in front of you: 4, U, 3, M. Which cards must you turn over in order to test the truth of the following proposition: If a card has a vowel on one side, then it has an even number on the other?" By the rules of logic, the statement is true if it cannot be refuted, which requires turning over the cards U and 3. However, the modal answer tends to be U and 4, although turning over 4 can only confirm the statement.

The primitives of the model in Sections 3 and 4 are the success rates p (as a result of taking the drug) and q (as a result of not taking the drug), as well as the specification of a choice procedure used by all patients. Our equilibrium concept is based on the following idea. Fix a firm's statement and a patients' choice procedure. Each fraction of patients who take the drug induces a distribution of cases and a probability that a new patient who samples from this distribution will buy the drug. An outcome is a fraction of patients who take the drug, which is equal to the probability that a new patient will take the drug. An outcome is stable if it is robust to small perturbations.

It will be shown that the notion of a stable outcome is well-defined in our model and therefore the firm's optimization problem is the following: Anticipating the stable outcome for each statement, the firm chooses one that maximizes the proportion of patients who use its drug.

The main contribution of the paper is the introduction of a new form of persuasion, in which a speaker makes a statement, which is not perceived as an informative signal, but rather is evaluated by individuals in a non-Bayesian way based on data they collect.

The model allows us to draw conclusions of the following types:

(i). The effectiveness of a statement "if A then B" depends on the procedure that individuals use to assess its validity. In particular, two logically equivalent statements, "if A then B" and "if not B, then not A", may have different persuasive appeal.

(ii). There are conditions under which positive statements are more (less) effective than negative statements.

(iii). There are conditions under which statements in which the antecedent is an action are more (less) persuasive than those in which the antecedent is a consequence.

(iv). Some procedures are better at reducing the likelihood of falling prey to imposters (i.e., when $p < q$), or increasing the chances of taking helpful actions ($p > q$).

The framework can also address questions regarding competing rhetoric (a debate) where two firms are trying to persuade a consumer, who will choose one of their products. Anticipating that patients will follow a choice procedure similar to those examined in Sections 3 and 4, each firm makes a statement that aims to maximize its market share. The question is then what type of statements will arise in equilibrium? Statements that defame the rival or statements that praise one's own product?

Section 5 models this situation as a strategic zero-sum game between two firms where one (the "superior" firm) offers a drug with a higher ex-ante chance of curing. We show that in equilibrium the support of each firm's strategy includes both a positive and a negative statement.

In particular, when the firms offer two "opposing" drug treatments (i.e., only one of the drugs can cure a patient, and its identity is independent across patients), the value of the game for the superior firm is just slightly below its success rate. If the inferior firm offers a useless drug (i.e. zero chance of it curing a patient) then in any equilibrium that firm always advertises using a negative statement about the rival and obtains a positive market share.

This paper relates to several strands of literature, and the final section provides an extensive discussion of this relation.

2 The basic model

A drug firm wishes to persuade a patient to use its product. The effect of the drug is binary: either it cures a patient or it does not. The truth (the state of the world) is such that a patient will be cured with probability $0 < p < 1$ if he takes the drug and with probability $0.5 \leq q < 1$ if he does not.[2] The effectiveness of the drug and the default is independent across patients. The probabilities p and q are known to the firm but not to the patients.

[2]The assumption that $q \geq 0.5$ is made only for simplicity of the analysis.

The firm makes a statement that if true would persuade the patient to take the drug. All statements are of the form $A \to B$ which is interpreted as: "if A is true then B must be true". We restrict attention to four statements: $[Y] \to [+]$, $[+] \to [Y]$, $[N] \to [-]$, and $[-] \to [N]$, where $[Y]$ and $[N]$ mean "taking the drug" and "not taking the drug", respectively, while $[+]$ and $[-]$ mean "being cured" and "not being cured", respectively. E.g., the statement $[+] \to [Y]$ is understood as "everyone who was cured took the drug".

The patients are *not* strategic - they don't take into account the firm's strategy when evaluating the truth of a statement. Furthermore, patients are not Bayesian and do not have a prior on their chance of being cured with the drug and without it. Rather, each patient uses a procedure that assesses the truth of the firm's statement that leads to either taking the drug or not.

Several variants of the model are considered, each of which is characterized by a different choice procedure that all patients use. Each procedure consists of observing past cases of patients who considered taking the drug, and applying a rule that evaluates the truth of the statement in light of the observed cases. A *case* has two components: whether the drug was taken [(Y) or (N)], and whether the patient was cured [(+) or (−)]. The cases are drawn from the population of cases. Given that a proportion b of the patients took the drug, the following table presents the case distribution:

case	content	frequency of case
Y^+	took the drug and were cured	$\alpha = bp$
Y^-	took the drug and were not cured	$\beta = b(1-p)$
N^+	did not take the drug and were cured	$\gamma = (1-b)q$
N^-	did not take the drug and were not cured	$\delta = (1-b)(1-q)$

Table 1: The cases and their frequencies.

Given a statement $A \to B$, we say that a case is a *confirmation* if both A and B are true and is a *refutation* if A is true and B is false. Given a distribution of cases and a statement s, denote the probability of a confirmation as $c_s(b)$ and the probability of a refutation as $r_s(b)$.

Given a choice procedure, each statement s and a case distribution induce a probability of acceptance denoted by $\Delta_s(b)$ (which depends on the procedure and the parameters p and q). For ease of exposition, we often omit the subscript s.

Given a procedure, an *outcome* of the statement s is a proportion b^* of patients who buy the drug, such that $\Delta_s(b^*) = b^*$. An outcome is *stable* if a small deviation from it triggers a push back to the outcome. Formally, an outcome b^* is stable if there is an $\epsilon > 0$ such that $\Delta_s(b) > b$ for any $b^* - \epsilon < b < b^*$ and $\Delta_s(b) < b$ for any $b^* + \epsilon > b > b^*$.

Each procedure is shown to induce a unique stable outcome for each statement s denoted by b^*_s. We investigate the firm's problem of finding a statement that leads to the highest stable outcome. Such a statement is referred to as *optimal*.

Comment: One question that arises is whether holding fixed the stable outcome b^* induced by an optimal statement s^*, the firm would prefer to deviate to a different statement s, i.e. $\Delta_s(b^*) > \Delta_{s^*}(b^*)$. If not, the optimal statement is said to be *locally stable*. In our setup, any optimal stable outcome is also locally stable. This is because for any statement, there exists a unique stable outcome, any acceptance function is analytical and its first derivative is equal to 1 at only finitely many points (and hence induces finitely many outcomes). Therefore, if s^* is optimal and there exists another statement s such that $\Delta_s(b^*) > \Delta_{s^*}(b^*) = b^*$, then the unique stable outcome of s is above b^*, either at a point where Δ_s crosses the main diagonal from above to below or at the point 1. But, this contradicts the optimality of s^*.

3 Finite samples

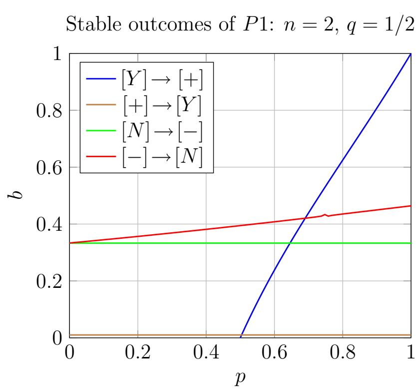
This section considers a procedure P1 in which a patient samples n cases and takes the drug after receiving statement s if the sample contains at least one confirmation and no refutations. The induced acceptance function is $\Delta_s(b) = (1 - r_s(b))^n - (1 - c_s(b) - r_s(b))^n$.

Notice that this procedure attributes significant weight to finding a refutation of the statement. This reflects the interpretation of a statement as a universal rule and thus even one refutation invalidates it, which seems to be intuitively appealing when the sample size is small. Note also that two logically equivalent statements (i.e., $A \to B$ and $-B \to -A$) do not induce the same acceptance functions.

Proposition 1: *Assume that patients use P1. Then, for any statement there is a unique stable outcome.*
For $n = 1$, the statements $[-] \to [N]$ and $[N] \to [-]$ are optimal.
For $n \geq 2$, there is $\hat{p}$ such that for $p > \hat{p}$ the optimal statement is $[Y] \to [+]$ and for $p < \hat{p}$ the optimal statement is $[-] \to [N]$.

According to this result, when the firm's drug is relatively effective (and a patient samples at least two cases) the firm's optimal statement argues that its product guarantees a cure. When the drug is relatively ineffective, the firm opts for the intimidating statement that all those who were not cured did not use the drug.

The following figure illustrates the proposition by depicting the stable outcomes (as a function of p) induced by each statement when $n = 2$ and $q = 1/2$.



Proof: In the case of $n = 1$, $\Delta_s(b) = c_s(b)$. The two positive statements are equivalent, $\Delta_{[Y]\to[+]}(b) = bp$ and the unique stable outcome is 0. The two negative statements are equivalent, $\Delta_{[N]\to[-]}(b) = (1-b)(1-q)$ and the unique stable outcome is $\frac{1-q}{2-q} > 0$. It follows that the two negative statements are optimal. Henceforth, we let $n > 1$.

In what follows we use two lemmas, which are proven in Appendix A:

Lemma 1: *For any statement s, each of the equations $\Delta'_s(b) = 0$ and $\Delta''_s(b) = 0$ has at most one solution in $[0,1]$.*

Lemma 2: *For any two statements s and s', which have the same confirmation case (refutation case), $\Delta_s(b) > \Delta_{s'}(b)$ iff $r_s(b) < r_{s'}(b)$ ($c_s(b) > c_{s'}(b)$) for all $b \in (0,1)$.*

We next show that each statement has a unique stable outcome and characterize it.

Lemma $[+] \to [Y]$: *The statement $[+] \to [Y]$ induces the unique stable outcome, $b = 0$.*

Proof: For this statement, $c_{[+] \mapsto [Y]}(b) = \alpha = bp$ and $r_{[+] \mapsto [Y]}(b) = \gamma = (1-b)q$. The acceptance function $\Delta_{[+] \mapsto [Y]}(b) = (1-\gamma)^n - (1-\gamma-\alpha)^n$ gets the value 0 at 0 and the value $1-(1-p)^n < 1$ at 1. The function Δ is convex since $\Delta''(b) = n(n-1)[q^2(1-\gamma)^{n-2} - (p-q)^2(1-\alpha-\gamma)^{n-2}] > 0$. Since $\Delta'(0) = n[q(1-q)^{n-1} + (p-q)(1-q)^{n-1}] = pn(1-q)^{n-1} < 1$, the statement induces a unique outcome $b = 0$ which is stable. $\square$

Lemma $[N] \to [-]$: *The statement $[N] \to [-]$ has a unique stable outcome which is below $\frac{(1-q)^2}{1-q+q^2}$.*

Proof: The acceptance function $\Delta_{[N] \mapsto [-]}(b) = (1-(1-b)q)^n - (b)^n$ satisfies $\Delta'(b) = nq(1-q+bq)^{n-1} - n(b)^{n-1}$, $\Delta(0) = (1-q)^n > 0$ and $\Delta(1) = 0$. Therefore, there must be an outcome which is stable.

By Lemma 1, there is at most one $b \in [0,1]$ where $\Delta'(b) = 0$. Since $\Delta'(0) > 0$ and $\Delta(1) = 0$ the function has a unique maximum point in $[0,1]$. It is attained when $q(1-\gamma)^{n-1} = b^{n-1}$ and at that point $\Delta(b) = \frac{1-\gamma}{q} b^{n-1} - b^n = (1-q)(1-(1-b)q)^{n-1}$. Since the last expression is decreasing in n, any outcome is at most the solution of $(1-q)(1-(1-b)q) = b$ which is $\frac{(1-q)^2}{1-q+q^2}$ and is at most $1/3$ for any $q \geq 1/2$.

In order to show that there is a unique stable outcome, we verify that in the interval $[0,1/3]$ the derivative $\Delta'(b) = nq(1-q+bq)^{n-1} - nb^{n-1} < 1$.

Fixing a value of b we look for $q \in [1/2, 1]$ that maximizes $\Delta'(b)$. If this maximizer is $q = 1/2$ then $\Delta'(b) \leq n\frac{1}{2}(2/3)^{n-1} < 1$ for all n. If this maximizer is $q = 1$, then $\Delta'(b) = 0$. Otherwise, there are two candidates for interior q that maximizes $\Delta'(b)$ (points where the derivative of $\Delta'(b)$ with respect to q vanishes): $q = \frac{1}{1-b}$ and $q = \frac{1}{n(1-b)}$. In the first case $\Delta'(b) \leq 0$. In the second case $\Delta'(b) = \frac{(1-\frac{1}{n})^{n-1}}{1-b} - nb^{n-1} \leq \frac{3}{2}(1-\frac{1}{n})^{n-1} < 1$ for all $n \geq 2$. $\square$

Lemma $[-]\rightarrow[N]$**:** *The statement* $[-]\rightarrow[N]$ *has a unique stable outcome.*

Proof: The function $\Delta_{[-]\mapsto[N]}(b) = (1-\beta)^n - (1-\beta-\delta)^n$ satisfies $\Delta(0) = 1 - q^n > 0$ and $\Delta(1) = 0$. It is decreasing in b (since the polynomial x^n is convex, $1-\beta$ is decreasing in b and δ is decreasing in b). Therefore, the statement induces a unique outcome which is stable. $\square$

Lemma $[Y]\rightarrow[+]$**:** *The statement* $[Y]\rightarrow[+]$ *induces a unique stable outcome which is* $b = 0$ *for* $n \leq 1/p$ *and is inside* $(0, 1]$ *for* $n > 1/p$*.*

Proof: The function $\Delta_{[Y]\mapsto[+]}(b) = (1-\beta)^n - (1-b)^n$ does not depend on q and satisfies $\Delta(0) = 0$, $\Delta(1) = p^n$ and $\Delta'(0) = -n(1-p) + n = np$. In addition, $\Delta''(b) = n(n-1)[(1-p)^2(1-\beta)^{n-2} - (1-b)^{n-2}]$ is negative at 0 and equal to zero at most once when $\frac{1-b}{1-\beta} = (1-p)^{2/(n-2)}$.

If $\Delta'(0) = np \leq 1$, the outcome $b = 0$ is stable. If there is an additional outcome b', then the first derivative is above 1 in a region below b', and since Δ' is decreasing near 0 there is an inflection point in $(0, b')$. Since $\Delta(1) < 1$ there is a region above b' where the derivative is below 1 and therefore there is an additional inflection point above b', contradicting Lemma 1.

For similar reasons, if $np > 1$ then the outcome $b = 0$ is unstable and there is exactly one positive outcome which is stable. $\square$

Remainder of the proof: We now show that $b^*_{[-]\mapsto[N]} \geq b^*_{[N]\mapsto[-]}$. The two statements share the confirmation case, and therefore, by Lemma 2, there is a unique intersection between the corresponding acceptance functions at $b' = \frac{q}{q+1-p}$ where the refutation rates are equal. By Lemma $[N] \rightarrow [-]$,

$b^*_{[N]\rightarrow[-]} \geq \frac{(1-q)^2}{1-q+q^2}$. Since the acceptance function $\Delta_{[-]\rightarrow[N]}$ is decreasing and $\Delta_{[N]\rightarrow[-]}(0) = (1-q)^n < \Delta_{[-]\rightarrow[N]}(0) = 1 - q^n$ it suffices to verify that $b' \geq \frac{(1-q)^2}{1-q+q^2}$ which is true for all $q \geq 1/2$.

Since $[+] \rightarrow [Y]$ induces the stable outcome 0, it remains to show that there is $\hat{p}$ such that $[-] \rightarrow [N]$ is optimal below it and $[Y] \rightarrow [+]$ is optimal above it. The statements share the refutation case (Y^-) and therefore by Lemma 2, $\Delta_{[-]\rightarrow[N]}(b) \geq \Delta_{[Y]\rightarrow[+]}(b)$ iff δ, the confirmation rate of $[-] \rightarrow [N]$, is at least as high as α, the confirmation rate of $[Y] \rightarrow [+]$, i.e., iff $b \leq \frac{1-q}{1+p-q}$.

Note also that the expression $\Delta_{[Y]\rightarrow[+]}(b) = (1 - b(1-p))^n - (1-b)^n$ is increasing in p and thus $b^*_{[Y]\rightarrow[+]}$ is increasing in p from 0 (at $p = 0$) to 1 (at $p = 1$).

Since $\Delta_{[-]\rightarrow[N]}$ is decreasing, the statement $[-] \rightarrow [N]$ is optimal for any p satisfying $b^*_{[Y]\rightarrow[+]} < \frac{1-q}{1+p-q}$ and the statement $[Y] \rightarrow [+]$ is optimal for any p satisfying $b^*_{[Y]\rightarrow[+]} > \frac{1-q}{1+p-q}$. Since the stable outcome of $[Y] \rightarrow [+]$ is an increasing function of p, there is $\hat{p}$ that splits the two regions. ∎

4 Searching until a relevant case is found

This section analyses three procedures in which a patient searches through past cases until he finds a "relevant" one. A key question is which cases are considered to be relevant for a statement of the form $A \rightarrow B$. Clearly, the confirmation and the refutation cases are relevant.

In procedure P2, a patient considers only those two cases to be relevant. In the other two procedures, a patient also makes inferences from an additional type of case. In P3, a case is relevant when the antecedent is false and the consequence is true. Such a case argues against taking the drug, because it is interpreted as evidence that the consequence is not necessarily caused by the antecedent. In P4, a case is relevant when both the antecedent and the consequence are false. Such a case argues in favor of taking the drug, because it is interpreted as evidence that the antecedent is indeed necessary to generate the consequence.

Table 2 summarizes the patients' behavior under each of the procedures.

$A \to B$	P2	P3	P4
$T\ T$	Y	Y	Y
$T\ F$	N	N	N
$F\ T$	$Pass$	N	$Pass$
$F\ F$	$Pass$	$Pass$	Y

Table 2: The sequential procedures.

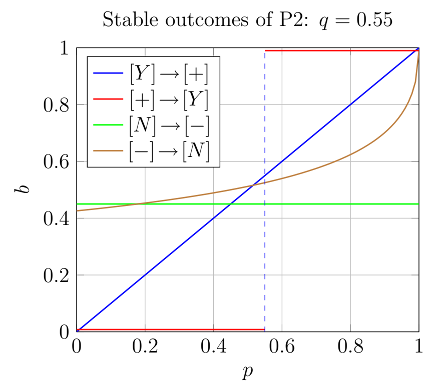
All acceptance functions in this section have the form $\Delta_s(b) = \frac{k_1 b + k_2}{k_3 b + k_4}$. Lemma 3, proven in Appendix A, establishes that such an acceptance function has a unique stable outcome:

Lemma 3: *If $\Delta_s(b) = \frac{k_1 b + k_2}{k_3 b + k_4}$, where both $k_1 b + k_2$ and $k_3 b + k_4$ receive values between 0 and 1 and $k_3 > 0$, then there is a unique stable outcome.*

In P2, a patient sequentially searches through cases and stops as soon as he observes either a confirmation or a refutation. The acceptance function for P2 is $\Delta_s(b) = \frac{c_s(b)}{c_s(b) + r_s(b)}$ and by Lemma 3, it has a unique stable outcome.

Proposition 2: *Assume that all patients use P2. The optimal statement and the unique stable outcome depend on the value of p as follows:*
Region I ($p \in [q, 1]$): $[+] \to [Y]$ and $b = 1$.
Region II ($p \in [\hat{p}, q]$): $[Y] \to [+]$ and $b = p$.
Region III ($p \in [1 - \frac{q^2}{1-q}, \hat{p}]$): $[-] \to [N]$ and $b = 1/(1 + \sqrt{\frac{1-p}{1-q}})$.
Region IV ($p \in [0, 1 - \frac{q^2}{1-q}]$): $[N] \to [-]$ and $b = 1 - q$.
where $\hat{p}$ is the unique solution of $\hat{p} = \frac{1}{1+\sqrt{\frac{1-\hat{p}}{1-q}}}$ in $(0, 1)$.

Under P2, every statement is optimal for some values of p and q. For relatively high p, the optimal statement is positive while for relatively low p it is negative. The warning that not taking the drug guarantees not being cured is optimal only for very low p. Its logically equivalent positive statement, namely that you will only be cured if you take the drug, is optimal only for $p \geq q$. The graph (taking $q = 0.55$) illustrates the proposition.



Proof: The proposition follows from the following observations:
The statement $[+] \to [Y]$ is optimal for $p > q$ (region I) since $\Delta_{[+]\to[Y]}(b) = \frac{bp}{bp+(1-b)q} > b$ in $(0,1)$, thus inducing the stable outcome 1. If $p < q$, then $\Delta_{[+]\to[Y]}(b) < b$ in $(0,1)$, thus inducing the stable outcome 0.

$\Delta_{[Y]\to[+]}(b) = p$, thus inducing the stable outcome p.

$\Delta_{[N]\to[-]}(b) = 1-q$, thus inducing the stable outcome $1-q$.

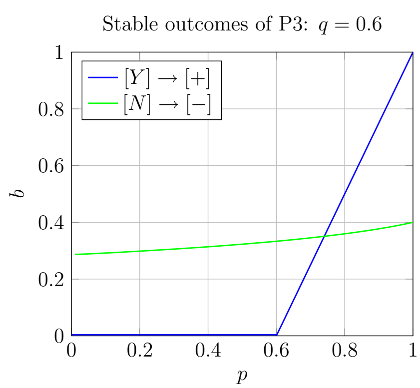
$\Delta_{[-]\to[N]}(b) = \frac{(1-b)(1-q)}{b(1-p)+(1-b)(1-q)}$ satisfies $\Delta(0) = 1$ and $\Delta(1) = 0$ and its stable outcome is $\frac{1}{1+\sqrt{\frac{1-p}{1-q}}}$.

The condition in region IV ($p \leq 1 - \frac{q^2}{1-q}$), is equivalent to the inequality $b^*_{[-]\to[N]}(p) \leq b^*_{[N]\to[-]}(p)$. In this region, $[N] \to [-]$ is also better than $[Y] \to [+]$ since $p \leq 1 - \frac{q^2}{1-q} \leq 1 - q$ (for any $q \geq 1/2$).

When $p \in [1 - \frac{q^2}{1-q}, q]$ (given that $q \geq 1/2$, this interval is not empty) the candidates for optimality are $[Y] \to [+]$ (with the outcome p) and $[N] \to [-]$ (with the outcome $\frac{1}{1+\sqrt{\frac{1-p}{1-q}}}$). In region III, the statement $[-] \to [N]$ is optimal up to $\hat{p} = \frac{1}{1+\sqrt{\frac{1-\hat{p}}{1-q}}}$ and $[Y] \to [+]$ is optimal in region II, above $\hat{p}$. ■

Under procedure P3 a patient stops searching and takes the drug as soon as he finds a confirmation. He stops and does not take the drug, as soon as he finds a refutation or a case where the antecedent of the statement is false but the consequence is true. The induced acceptance function for a statement s is $\Delta_s(b) = \frac{c_s(b)}{1-n_s(b)}$, where $n_s(b)$ is the probability of a case in which both the antecedent and the consequence of s are false. By Lemma 3, it induces a unique stable outcome. Under this procedure, there is no difference between any two statements $A \to B$ and $B \to A$ and hence it is sufficient to compare only $[Y] \to [+]$ and $[N] \to [-]$. The next proposition, proven in Appendix A, characterizes the optimal statement for different values of p :

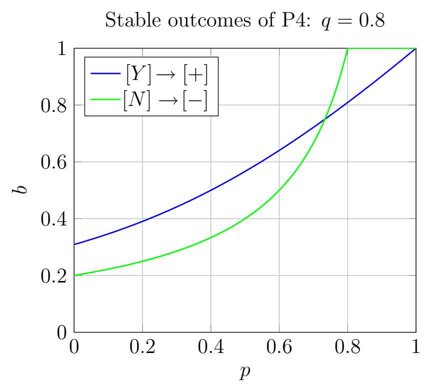
Proposition 3: *If all patients use procedure P3, then there is $\hat{p}$ such that for any $p < \hat{p}$ the statement $[N] \to [-]$ is optimal and for $p > \hat{p}$ the statement $[Y] \to [+]$ is optimal. The stable outcomes of the optimal statement are interior for all $p < 1$.*



Once again the positive statement is optimal in the top range of p and the negative statement is optimal in the bottom range. The proposition is illustrated in the figure for $q = 0.6$.

Finally, under procedure P4 a patient stops searching and takes the drug as soon as he finds either a confirmation of the statement or a case in which both the antecedent and consequence are false. He stops searching and does not take the drug as soon as he finds a refutation of the statement. Under P4, the acceptance function given the statement s is $\Delta_s(b) = \frac{c_s(b)+n_s(b)}{c_s(b)+n_s(b)+r_s(b)}$. By Lemma 3, it has a unique stable outcome. Under this procedure there is no difference between any two logically equivalent statements $A \to B$ and $\neg B \to \neg A$, and we are left to compare $[Y] \to [+]$ and $[N] \to [-]$.

The next proposition (proven in Appendix A and illustrated in the figure for $q = 0.8$) shows that in contrast to the result regarding the two previous procedures, $[Y] \to [+]$ is optimal for relatively low p while the statement $[N] \to [-]$ is optimal for high p.



Proposition 4: *Assume that all patients use procedure P4. If* $q < \frac{\sqrt{5}-1}{2}$, *then* $[N] \to [-]$ *is optimal for all* p *(if* $p > q$ *the stable outcome is* 1 *and otherwise it is* $\frac{1-q}{1-p}$*). If* $\frac{\sqrt{5}-1}{2} < q$*, then there is* $\hat{p}$ *such that* $[Y] \to [+]$ *is optimal when* $p < \hat{p}$ *and* $[N] \to [-]$ *is optimal when* $p > \hat{p}$*.*

Comment: We conclude the section by comparing the sequential search procedures from the point of view of a patient. When $p > q$, procedure $P2$ is the best since it leads to the desired stable outcome of 1. When $p < 1/2 \leq q$, the best procedure for the patients minimizes the likelihood of taking the drug, which in this case is worse than not taking it. This procedure is identified in Proposition 5 which is proven in Appendix A:

Proposition 5: *When* $p < 1/2$*, the* $P3$ *procedure is better for a patient than* $P2$ *and* $P4$*.*

5 Dueling statements

Two firms, 1 and 2, compete over a population of patients by offering drugs with success rates of $p_1 > p_2$, respectively. These rates are known only to the firms. Eventually, every patient in the population buys one of the drugs. Each firm simultaneously makes one of the four statements described in the previous sections, although they now have a different interpretation: Y means "buying from the firm that made the statement" and N means "buying from its rival". Thus, for example, the statement $[N] \to [-]$ is interpreted as "anyone who does not buy from our firm will not be cured".

We assume that all patients use the following procedure: Sample until you find either a confirmation or a refutation of one of the statements; buy from a firm if you sampled a confirmation of that firm's statement or a refutation of its rival's.

Firm i's payoff from a pair of statements s_1, s_2, denoted by $u_i(s_1, s_2)$, is calculated as follows: Let b denote the proportion of the population that buys from firm 1. Denote by $W_i(s_1, s_2, b)$ the probability that s_i is confirmed or s_{-i} is refuted in a single draw. Define $\Delta_{(s_1,s_2)}(b)$ to be the acceptance function of firm 1's drug when the firms make the statements s_1 and s_2 and b is the fraction of the population that buys the drug from firm 1. Given the patients' procedure, $\Delta_{(s_1,s_2)}(b) = \frac{W_1(s_1,s_2,b)}{W_1(s_1,s_2,b)+W_2(s_1,s_2,b)}$. Denote by $b^*(s_1, s_2)$ the stable solution to the equation $\Delta_{(s_1,s_2)}(b) = b$, which will be shown to always exist and to be unique. Finally, let $u_1(s_1, s_2) = b^*(s_1, s_2)$ and $u_2(s_1, s_2) = 1 - b^*(s_1, s_2)$. This completes the description of the interaction between the firms as a strategic zero-sum game.

As in the single-firm analysis, the above definition of payoffs adopts an *ex-ante* view: When firm i compares a statement s to a statement s' given that its rival makes the statement s_j, it *forecasts* the acceptance probabilities under each pair of statements.

Denote by 1^+ the case in which firm 1's statement is confirmed and by 1^- the case in which it is refuted and similarly for firm 2. Thus, for example, if 1 makes the statement $[2] \to [-]$ and 2 makes the statement $[+] \to [2]$, then 1 wins when a patient observes 2^- (confirming 1's statement) or 1^+ (refuting 2's statement). It loses only in case 2^+ (refuting 1's statement and confirming 2's). In order to calculate the payoff matrix we construct a table in which the entry in row s_1 and column s_2 contains on its left, the cases in which 1 is chosen, and on its right, the cases in which 2 is chosen.

	$[2]\rightarrow[+]$	$[+]\rightarrow[2]$	$[1]\rightarrow[-]$	$[-]\rightarrow[1]$
$[1]\rightarrow[+]$	$1^+2^-/1^-2^+$	$1^+/1^-2^+$	$1^+/1^-$	$1^+2^-/1^-$
$[+]\rightarrow[1]$	$1^+2^-/2^+$	$1^+/2^+$	$1^+/2^+1^-$	$1^+2^-/2^+1^-$
$[2]\rightarrow[-]$	$2^-/2^+$	$2^-1^+/2^+$	$2^-1^+/2^+1^-$	$2^-/2^+1^-$
$[-]\rightarrow[2]$	$2^-/1^-2^+$	$2^-1^+/1^-2^+$	$2^-1^+/1^-$	$2^-/1^-$

Table 3: The winning/losing cases.

Note that two pairs of statements induce the same acceptance function if they lead to the same sets of winning and losing cases (e.g. the four pairs of statements for which 1 wins in cases 1^+ and 2^- and loses in the other two cases). Table 5 in Appendix B presents the corresponding acceptance function and induced unique stable outcome for each entry in Table 3. Thus, the payoff matrix is as follows:

		b_1	b_2	b_3	b_4
		$[2]\rightarrow[+]$	$[+]\rightarrow[2]$	$[1]\rightarrow[-]$	$[-]\rightarrow[1]$
a_1	$[1]\rightarrow[+]$	T_4	T_3	p_1	T_1
a_2	$[+]\rightarrow[1]$	1	1	T_3	T_4
a_3	$[2]\rightarrow[-]$	$1-p_2$	1	T_4	T_2
a_4	$[-]\rightarrow[2]$	T_2	T_4	T_1	T_0

Table 4: The payoff matrix.

The exact expressions for T_0, T_1, T_2, T_3 and T_4 are shown in Table 5 in Appendix B. Note that $T_1 > T_4 > T_2, T_3$; $\quad T_1 > p_1 > T_3$; $1 - p_2 > T_2$ and $T_4 > T_0 > T_2$.

The game does not have an equilibrium in which a firm plays a pure strategy. Given that $p_1 > p_2$ the value of the game is greater than $1/2$. This is because for any strategy of firm 2, firm 1 can achieve a market share larger than $1/2$ by mimicking firm 2's strategy (thus obtaining an expected payoff that is a convex combination of the values T_0, $(T_1+T_2)/2$, T_4, $(p_1+1-p_2)/2$, and $(1+T_3)/2$, which are all above $1/2$).

Our main observation about the game is that in every equilibrium each firm uses both a positive and a negative statement.

Proposition 6: *In every equilibrium, the support of each firm's strategy includes at least one positive and one negative statement. Furthermore, the support of the strategy of firm 2 (the firm with the inferior product) includes* $[1] \rightarrow [-]$.

Proof: Let (σ_1, σ_2) be a mixed-strategy Nash equilibrium. Assume first that the support of σ_2 does not include $b_3 = [1] \rightarrow [-]$. Then, firm 1's statements a_3 and a_4 are dominated by a_2. But then b_4 is dominated (by b_3) and therefore a_2 dominates a_1, a contradiction to any equilibrium not involving a pure strategy.

Second, assume that the support of σ_1 does not include a positive statement. Then b_3 is dominated by b_4, a contradiction of the previous step.

Third, assume that the support of σ_1 does not include a negative statement. Then, b_1 and b_4 are dominated by b_2 and b_3, respectively. But then a_1 is dominated by a_4, a contradiction.

Finally, assume that σ_2's support contains only the two negative statements. Then, a_2 and a_3 are dominated by a_1 and a_4, respectively. But now b_3 is dominated by b_2, a contradiction. ■

More can be said about the equilibria of this game in two special cases. The first is $p_1 + p_2 = 1$, which can be interpreted as a market in which two firms offer opposite drug treatments, only one of which can cure a patient. The identity of the useful drug is independent across patients and the probability that it is firm i's drug is p_i.

Proposition 7: *Assume that $p_1 = p > p_2 = 1 - p$. (i) In any equilibrium the support of firm 1's strategy includes both of its positive statements while the support of firm 2's strategy includes both of its negative statements. (ii) The value of the game is below p but above $p - 0.015$. (iii) Firm 2 assigns probability higher than p to the negative statements.*

The result that the market share of the superior firm is almost p can be compared to those of two other cases. Suppose firm 1 is the only firm in the market, patients search until they find a previous patient who bought the drug, and then decide to also buy the drug if and only if the sampled patient was cured. In this case, firm 1's share will be p, and hence the presence of an inferior competitor has only a small impact on its market share.

If, on the other hand, the market also includes a second inferior firm that doesn't advertise, then by Proposition 2, firm 1 will capture the entire market. The reason that the presence of another firm (which doesn't advertise) boosts firm 1's market share is as follows: When there is no other firm, encountering a patient who was not cured is bad news for firm 1 since it is the only firm offering a drug. However, when there is another firm that sells the drug, sampling a patient who was not cured can be good news for firm 1 since the sampled patient might have bought the drug from firm 2.

Propositions 6 and 7 imply that the support of an equilibrium strategy contains at least three statements. Calculations show that for relatively low (close to 0.5) and relatively high (close to 1) values of p there is an equilibrium with full supports. Such equilibria have an interesting structure. Both firms assign the same probability to the positive and negative statements that a consequence implies an action, i.e., the probabilities that firm 1 assigns to $[+] \to [1]$ and $[-] \to [2]$ are equal to those which firm 2 assigns to $[+] \to [2]$ and $[-] \to [1]$, respectively. However, they swap the probabilities of the statements that an action implies a consequence, i.e., the probabilities that firm 1 assigns to $[1] \to [+]$ and $[2] \to [-]$ are equal to those which firm 2 assigns to $[1] \to [-]$ and $[2] \to [+]$, respectively. The reason for this is that if σ_1's support includes all four statements then it must be that the expected payoffs of all statements given $\sigma_2 = (\alpha, \beta, \gamma, \delta)$ are equal. However, it is easy to verify that $\sigma_1 = (\gamma, \beta, \alpha, \delta)$ then makes firm 2 indifferent between all statements and thus it is a maxmin strategy for firm 1.

Proof of Proposition 7:

(i) Consider the strategy $\sigma_2 = (\frac{T_1 - T_3}{1 - T_2 + T_1 - T_3}, 0, \frac{1 - T_2}{1 - T_2 + T_1 - T_3}, 0)$. Firm 1's expected payoff from a_2 and a_4 is equal to $\frac{T_1 - T_3}{1 - T_2 + T_1 - T_3} \cdot 1 + \frac{1 - T_2}{1 - T_2 + T_1 - T_3} \cdot T_3$, which is less than p iff $(T_1 - T_3)(1 - p) < (1 - T_2)(p - T_3)$ which holds when $1/2 < p < 1$. Therefore, firm 2 can guarantee that firm 1's share will not exceed p.

Next, consider the strategy $\sigma_1 = (p, 0, 1 - p, 0)$. Since $pT_3 + (1 - p) = p$ firm 1's expected share given the statements b_1, b_2 or b_3 is p. Firm 1's share if firm 2 uses b_4 is $pT_1 + (1 - p)T_2$, which is above $p - 0.0154$.

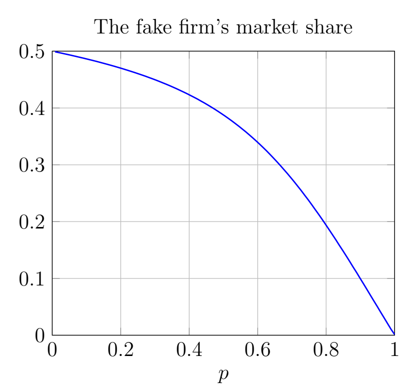
(ii) If 2 assigns probability lower than p to the negative statements, then firm 1's expected payoff from a_2 is more than $(1 - p) + pT_3 = p$.

(iii) See Appendix B. ■

The second special case is one in which only firm 1 offers an effective drug ($p_1 > 0$ and $p_2 = 0$). The question in this case is whether firm 2 can guarantee itself a positive market share despite the fact that it offers a "fake" drug, and if so, which type of statements enables it to do so.

Proposition 8: *When $p_1 = p > p_2 = 0$, there is a unique mixed-strategy equilibrium. In that equilibrium, the support of firm 1's strategy contains only the statements* $[+] \to [1]$ *and* $[2] \to [-]$*; the support of firm 2's strategy contains only* $[1] \to [-]$ *and* $[-] \to [1]$*; and the value of the game is* $\frac{(T_1)^2 - pT_2}{2T_1 - p - T_2} > p$.

Thus, the firm that offers a fake drug uses only negative statements about its rival and obtains a positive share of the market. The graph plots the market share obtained by the fake firm for different values of p:



Proof: In this case, $T_0 = T_2$, $T_3 = p$ and $T_4 = T_1$. Elimination of weakly dominated strategies leaves firm 1 with $[+] \to [1]$ and $[2] \to [-]$, and firm 2 with the two negative statements. A routine calculation yields the value of the game, which is above p since $T_1 > p$. ■

6 Related literature

We conclude by comparing this paper to related strands of the literature in which a speaker tries to persuade a listener to take some action.

Cheap talk. Although our model features cheap-talk statements by a speaker, it differs from the standard cheap talk framework (see Crawford and Sobel, 1982). All that matters in that framework is the correlation between messages and the state: The content and framing of the message per se are meaningless. In contrast, the content of a statement made by our speaker is taken and assessed by the listener literally and different framings of a statement can be differentially effective. While the speaker's statement may be correlated with his knowledge, our listener is non-strategic and does not draw inferences from that correlation.

Non-Bayesian persuasion In some persuasion models, the listener is not Bayesian either because he commits how to respond to any information received from the speaker (e.g., Glazer and Rubinstein, 2004), or because he makes systematic mistakes in updating his beliefs given the speaker's message (e.g., de Clippel and Zhang, 2022).

Bayesian persuasion. In this approach (see Kamenica and Gentzkow, 2011), a speaker publicly commits ex-ante to a Blackwell experiment, and a Bayesian listener responds to the experiment's outcome. In contrast, our listener is not engaged in belief updating. Instead, he samples past cases and evaluates the speaker's statement in light of his sample, which stochastically depends on the equilibrium behavior of listeners.

Narratives. Our speaker makes a statement that relates an outcome to an action. It can be interpreted as making a claim about a *causal* relation between taking the drug and being cured. In this sense our paper is related to recent studies on narratives as causal models. Under this view (see Eliaz and Spiegler, 2020 and Eliaz, Spiegler and Weiss, 2021), a speaker, who knows the true prior distribution over several variables, announces a causal relation between a subset of these variables. The listener accepts this relation and forms a belief by applying a Bayesian method to infinite data on the realizations of all the variables, data which (as in our framework) is affected by the listeners' response to the narrative. In contrast to our model the listeners *are* Bayesian, they do *not* test the validity of the narrative, and they use an infinite amount of data.

Advertising. Advertising is traditionally modeled as an action taken by a firm to draw the consumer's attention to its product, change his utility from its product or affect the consumer's beliefs (see Bagwell, 2007). In contrast, advertising in our model is a qualitative cheap talk statement about the product's quality, which is assessed by a consumer using data on other consumers' equilibrium behavior.

Sampling. The assumption that a listener samples cases from the equilibrium distribution builds on Osborne and Rubinstein (1998). Our result in Section 5, that in equilibrium a positive proportion of patients buy a drug even when it is useless, is reminiscent of Spiegler (2006). He showed that when firms compete in prices and consumers use a sampling procedure, firms earn positive profits in equilibrium even though their product is worse than a costless default. See also Bianchi and Jehiel (2015) who study investors who pay attention only to a randomly drawn sample from firms' financial reports.

7 References

Bagwell, Kyle (2007). "The economic analysis of advertising", in *Handbook of Industrial Organization Volume 3* (Mark Armstrong, Robert H. Porter, editors). Elsevier.

Bianchi, Milo, and Philippe Jehiel (2015). "Financial reporting and market efficiency with extrapolative investors." *Journal of Economic Theory* 157: 842-878,

Crawford, Vincent P., and Joel Sobel (1982). "Strategic information transmission." *Econometrica* 50: 1431-1451.

De Clippel, Geoffroy, and Xu Zhang (2022). "Non-bayesian persuasion." *Journal of Political Economy*, 130: 2594-2642.

Eliaz, Kfir, and Ran Spiegler (2020). "A model of competing narratives." *American Economic Review* 110: 3786-3816.

Eliaz, Kfir, Ran Spiegler, and Yair Weiss (2021). "Cheating with models." *American Economic Review: Insights* 3: 417-434.

Glazer, Jacob, and Ariel Rubinstein (2004). "On optimal rules of persuasion." *Econometrica* 72: 1715-1736.

Kamenica, Emir, and Matthew Gentzkow (2011). "Bayesian persuasion." *American Economic Review* 101: 2590-2615.

Klayman, Joshua, and Young-Won Ha (1987). "Confirmation, disconfirmation, and information in hypothesis testing." *Psychological Review* 94: 211-228.

Osborne, Martin J., and Ariel Rubinstein (1998). "Games with procedurally rational players." *American Economic Review* 88: 834-847.

Spiegler, Ran (2006). "The market for quacks." *The Review of Economic Studies* 73: 1113-1131.

Wason, Peter Cathcart (1960). "On the failure to eliminate hypotheses in a conceptual task." *Quarterly Journal of Experimental Psychology* 12: 129-140.

Wason, Peter Cathcart, and Philip Nicholas Johnson-Laird (1972). *Psychology of Reasoning: Structure and Content*. Harvard University Press.

Appendix A (to sections 3 and 4)

Lemma 1: *For any statement s, each of the equations $\Delta'_s(b) = 0$ and $\Delta''_s(b) = 0$ has at most one solution in $[0, 1]$.*

Proof: Since c_s and r_s are linear functions of b then $\Delta_s(b)$ has the form $\Delta_s(b) = (k_1 b + k_2)^n - (k_3 b + k_4)^n$, where: (i) $k_1 b + k_2 \geq k_3 b + k_4$ for all b (since $\Delta_s(b) \geq 0$), (ii) $k_1 \neq k_3$ (since $c_s(b)$ is not constant) and (iii) $k_1 b + k_2 > 0$ for all b (since $r_s(b) < 1$ for all b).

The equation $\Delta'_s(b) = n[k_1(k_1 b + k_2)^{n-1} - k_3(k_3 b + k_4)^{n-1}] = 0$ can hold only when $k_3 b + k_4 > 0$ in which case the equation is equivalent to the linear equation $\frac{k_1 b + k_2}{k_3 b + k_4} = (\frac{k_3}{k_1})^{1/(n-1)}$. For even n, there is a unique root which is positive and therefore there is a unique b that solves the linear equation (which might be outside the interval $[0, 1]$). For odd n, there are two roots, but the negative root induces a solution that is negative and therefore only the positive root induces an equation with a solution that might be inside $[0, 1]$.

Similarly, $\Delta''_s(b) = n(n-1)[k_1^2(k_1 b + k_2)^{n-2} - k_3^2(k_3 b + k_4)^{n-2}] = 0$ holds when $\frac{k_1 b + k_2}{k_3 b + k_4} = (\frac{k_3^2}{k_1^2})^{1/(n-2)}$ and thus can be true in $[0, 1]$ at most once. $\square$

Lemma 2: *For any two statements s and s', which have the same confirmation case (refutation case), we have $\Delta_s(b) > \Delta_{s'}(b)$ iff $r_s(b) < r_{s'}(b)$ ($c_s(b) > c_{s'}(b)$) for all $b \in (0, 1)$.*

Proof: If the statements share the same refutation case then the comparison between the acceptance functions is simply between the confirmation rates. If they share the same confirmation case, then for each $b \in (0, 1)$ we have $c_s(b) = c_{s'}(b) > 0$, and by the convexity of the polynomial x^n, the comparison $\Delta_s(b) > \Delta_{s'}(b)$ is equivalent to $r_s(b) < r_{s'}(b)$. $\square$

Lemma 3: *If $\Delta_s(b) = \frac{k_1 b + k_2}{k_3 b + k_4}$, where both $k_1 b + k_2$ and $k_3 b + k_4$ receive values between 0 and 1 and $k_3 > 0$, then there is a unique stable outcome.*

Proof: The equation $\Delta_s'(b) = 1$ is equivalent to $k_1k_4 - k_2k_3 - (k_3b + k_4)^2 = 0$. Since $k_3b + k_4$ is positive the equation has at most one solution and therefore there is at most one point where $\Delta_s'(b) = 1$.

The acceptance function is continuous and therefore there is at least one stable outcome: If the graph of the function lies above the main diagonal, then the point $b = 1$ is stable; if it lies below it, then the point $b = 0$ is stable and otherwise it has at least one stable outcome when it crosses the main diagonal from above.

Assume by contradiction that there are two stable outcomes $b_1 < b_2$. Then $\Delta_s'(b_1) < 1$ and $\Delta_s'(b_2) < 1$. In order to reach b_2 from above the main diagonal, there must be a point $b_1 < b_3 < b_2$ at which $\Delta_s'(b_3) > 1$. By the continuity of Δ_s', there are two points where $\Delta_s'(b) = 1$, a contradiction. □

Proposition 3: *If all patients use procedure P3, then there is $\hat{p}$ such that for any $p < \hat{p}$ the statement $[N] \to [-]$ is optimal and for $p > \hat{p}$ the statement $[Y] \to [+]$ is optimal. The stable outcomes of the optimal statements are interior for all $p < 1$.*

Proof: The function $\Delta_{[Y]\mapsto[+]}(b) = \frac{\alpha}{1-\delta} = \frac{bp}{b+(1-b)q}$ satisfies $\Delta(0) = 0$, $\Delta(1) = p$ and $\Delta'(0) = p/q$. For $p < q$, the stable outcome is $b = 0$ and for $p \geq q$ it is $b^*_{[Y]\mapsto[+]}(p) = \frac{p-q}{1-q}$.

The function $\Delta_{[N]\to[-]}(b) = \frac{\delta}{1-\alpha} = \frac{(1-b)(1-q)}{1-bp}$ satisfies $\Delta(0) = 1 - q$ and $\Delta(1) = 0$ and therefore it has an interior stable outcome. Thus, for $p \leq q$ the statement $[N] \to [-]$ is optimal ($b^*_{[N]\to[-]}(p) = \frac{2-q-\sqrt{(2-q)^2-4p(1-q)}}{2p}$).

When $p > q$, clearly, $b^*_{[N]\to[-]}(q) > b^*_{[Y]\mapsto[+]}(q)$ and $b^*_{[N]\to[-]}(1) < b^*_{[Y]\mapsto[+]}(1)$. The equation $b^*_{[N]\to[-]}(p) = b^*_{[Y]\mapsto[+]}(p)$ requires that $\Delta_{[N]\mapsto[-]}$ and $\Delta_{[Y]\mapsto[+]}$ intersect at $b_0 = \frac{p-q}{1-q}$. Since $\delta(b_0)(1 - \delta(b_0)) = (1-p)p$ and $\alpha(b_o)(1 - \alpha(b_0)) = p\frac{(p-q)(1-q-p^2+pq)}{(1-q)^2}$, this occurs only at p which solves the equation $(p-q)(1+p-q) = (1-q)^2$ i.e. at $\hat{p} = \frac{\sqrt{(1-2q)^2+4(1-q)}+2q-1}{2}$. The statement $[N] \to [-]$ is optimal for $p < \hat{p}$ and $[Y] \to [+]$ is optimal for $p > \hat{p}$. ■

Proposition 4: *Assume that all patients use procedure P4. If* $q < \frac{\sqrt{5}-1}{2}$*, then* $[N] \to [-]$ *is optimal for all p (if* $p > q$ *the stable outcome is* 1 *and otherwise it is* $\frac{1-q}{1-p}$*). If* $\frac{\sqrt{5}-1}{2} < q$*, then there is* $\hat{p}$ *such that* $[Y] \to [+]$ *is optimal when* $p < \hat{p}$ *and* $[N] \to [-]$ *is optimal when* $p > \hat{p}$*.*

Proof: The function $\Delta_{[N]\to[-]}(b) = \frac{\alpha+\delta}{\alpha+\delta+\gamma}$ satisfies that $\Delta(0) = 1 - q$ and $\Delta(1) = 1$. For $p > q$, we have $\Delta(b) > b$ for all $b < 1$ and the stable outcome is $b = 1$. For $p < q$, the point $b = 1$ is unstable and therefore $b^*_{[N]\to[-]} = \frac{1-q}{1-p}$.

The function $\Delta_{[Y]\to[+]}(b) = \frac{\alpha+\delta}{1-\gamma}$ satisfies $\Delta(0) = 1 > 1 - q$, $\Delta(1) = p < 1$ and its stable outcome is $b^*_{[Y]\to[+]} = \frac{2q+p-2+\sqrt{(2-2q-p)^2+4q(1-q)}}{2q}$.

To compare the two statements when $p < q$, note that $\Delta_{[N]\to[-]}(b) = \Delta_{[Y]\to[+]}(b)$ only when $b = \frac{q}{q+1-p}$. The statement $[N] \to [-]$ is optimal iff $b^*_{[N]\to[-]} = \frac{1-q}{1-p} \geq \frac{q}{q+1-p}$. For $q \leq (\sqrt{5} - 1)/2$ it holds for all $p \leq q$ and otherwise it holds in the interval $p \in [\frac{q^2+q-1}{2q-1}, q]$. ■

Proposition 5: *When* $p < 1/2$ *the P3 procedure is weakly better for the patient than P2 and P4.*

Proof: We proceed in several steps. Denote by $\Delta_i(b)$ the acceptance function of the firm's optimal statement given procedure Pi and by b^*_i the unique stable outcome induced by the firm's optimal statement.

Step 1: $\Delta_4(b) = \frac{(1-b)(1-q)}{1-bp}$.

Since $p < 1/2 \leq q$ the inequality $p < \frac{\sqrt{(1-2q)^2+4(1-q)}+(2q-1)}{2}$ holds (since the RHS of the inequality is above $1/2$) and therefore by Proposition 3 the firm's optimal statement given $P3$ is $[N] \to [-]$ and therefore $\Delta_4(b) = \frac{(1-b)(1-q)}{1-bp}$.

Step 2: *P3 induces a weakly lower stable outcome than P2.*

It suffices to show that $\Delta_4(b) \leq \Delta_3(b)$ for all b. By Proposition 2:
If $p \in [0, 1 - \frac{q^2}{1-q}]$, then $\Delta_3(b) = 1 - q \geq \frac{1-b}{1-bp}(1 - q) = \Delta_4(b)$.
If $p \in \left[1 - \frac{q^2}{1-q}, \hat{p}\right]$, then $\Delta_3(b) = \frac{(1-b)(1-q)}{b(1-p)+(1-b)(1-q)} \geq \Delta_4(b) = \frac{(1-b)(1-q)}{1-bp}$ iff $1 - bp \geq b - bp + (1 - b)(1 - q)$, which is true.

Finally, if $p \in [\hat{p}, q]$, then $p \geq \hat{p} = \frac{1}{1+\sqrt{\frac{1-\hat{p}}{1-q}}} = \sqrt{1-q}\frac{\sqrt{(5-q)}-\sqrt{(1-q)}}{2} > 1-q$. Hence, $\Delta_3(b) = p > (1-q) \geq \frac{1-b}{1-bp}(1-q) = \Delta_4(b)$.

Step 3: *$P3$ induces a lower stable outcome than $P4$ when $[Y] \rightarrow [+]$ is the firm's optimal statement under $P4$.*

Assume that $[Y] \rightarrow [+]$ is the optimal statement under $P4$. Then by Proposition 4, $\Delta_5(b) = \frac{bp+(1-b)(1-q)}{1-(1-b)q}$ whereas $\Delta_4(b) = \frac{(1-b)(1-q)}{1-bp}$. Hence, $\Delta_4(b) < \Delta_5(b)$ for all $b < \frac{q}{q+p}$. It therefore suffices to show that $b_5^* < \frac{q}{q+p}$. Some tedious algebra indeed verifies this inequality for all $q \geq 1/2 > p$.

Step 4: *$P3$ induces a lower stable outcome than $P4$ when $[N] \rightarrow [-]$ is the firm's optimal statement under $P4$.*

By Proposition 4, when $[N] \rightarrow [-]$ is optimal given $P4$, we have $\Delta_5(b) = \frac{bp+(1-b)(1-q)}{1-b(1-p)} > \Delta_4(b) = \frac{(1-b)(1-q)}{1-bp}$, since given $p < 1/2$ we have both $bp + (1-b)(1-q) > (1-b)(1-q)$ and $1 - b(1-p) < 1 - bp$. ■

Appendix B (to section 5)

combination	$\Delta_{(s_1,s_2)}$	The Stable Outcome
$1^+/2^+$	$\frac{bp_1}{bp_1+(1-b)p_2}$	1
$2^-/2^+$	$1-p_2$	$1-p_2$
$1^+/1^-$	p_1	p_1
$2^-/1^-$	$\frac{(1-b)(1-p_2)}{b(1-p_1)+(1-b)(1-p_2)}$	$T_0=\frac{\sqrt{1-p_2}}{\sqrt{1-Tp_1}+\sqrt{1-p_2}}$
$2^-1^+/2^+$	$\frac{bp_1+(1-b)(1-p_2)}{1-b(1-p_1)}$	1
$1^+2^-/1^-$	$\frac{bp_1+(1-b)(1-p_2)}{1-(1-b)p_2}$	$T_1=\frac{2p_2+p_1-2+\sqrt{(2-p_1)^2-4p_2(1-p_1)}}{2p_2}$
$2^-/2^+1^-$	$\frac{(1-b)(1-p_2)}{1-bp_1}$	$T_2=\frac{2-p_2-\sqrt{(2-p_2)^2-4p_1(1-p_2)}}{2p_1}$
$1^+/2^+1^-$	$\frac{bp_1}{1-(1-b)(1-p_2)}$	$T_3=\frac{p_1-p_2}{1-p_2}$
$2^-1^+/1^-2^+$	$bp_1+(1-b)(1-p_2)$	$T_4=\frac{1-p_2}{2-p_1-p_2}$

Table 5: The acceptance functions and the unique stable outcome for each acceptance/rejection combination, where $T_1 > T_4, p_1$; $T_4, 1-p_2 > T_2$; p_1, $T_4 > T_3$; $T_1 > T_4 > T_2, T_3$.

		b_1	b_2	b_3	b_4
		$[2]\to[+]$	$[+]\to[2]$	$[1]\to[-]$	$[-]\to[1]$
a_1	$[1]\to[+]$	p	T_3	p	T_1
a_2	$[+]\to[1]$	1	1	T_3	p
a_3	$[2]\to[-]$	p	1	p	T_2
a_4	$[-]\to[2]$	T_2	p	T_1	T_0

Table 6: The payoff matrix for the case $p_1 = p = 1-p_2$, where $T_0 = \frac{p}{p+\sqrt{p(1-p)}}$; $T_1 = \frac{-p+\sqrt{(2-p)^2-4(1-p)^2}}{2(1-p)} > p > T_2 = \frac{1+p-\sqrt{(1+p)^2-4p^2}}{2p}$; $p > T_3 = \frac{2p-1}{p}$ and $p > T_0 > T_2$.

Proposition 7: *Assume that $p_1 = p > p_2 = 1-p$. (iii) In any equilibrium, the support of firm 1's strategy includes both of its positive statements while the support of firm 2's strategy includes both of its negative statements.*

Proof: Let (σ_1, σ_2) be an equilibrium:

Step 1: *It is impossible that the support of σ_1 includes only a_1 and a_3.*
Assume the contrary. We obtain a contradiction by showing that either b_2 or b_4 is superior to b_3 (which by Proposition 6 is in the support of σ_2). That is, for every α, either $\alpha \cdot T_3 + (1-\alpha) \cdot 1 < p$ or $\alpha \cdot T_1 + (1-\alpha) \cdot T_2 < p$. If the first inequality does not hold, that is, if $1 - p \geq \alpha(1 - T_3)$, then we need to show that $\alpha \cdot T_1 + (1-\alpha) \cdot T_2 < p$. It suffices to show that $\frac{1-p}{1-T_3}(T_1 - T_2) < p - T_2$, i.e., $p(1 - T_1) > T_2(1 - p)$ which holds for all $1 > p > 1/2$.

Step 2: *The support of σ_1 includes a_2.*
If not, b_3 is dominated by b_1 since by Step 1, σ_1's support includes a_4.

Step 3: *The support of σ_2 includes b_4.*
Assume the contrary. There are two possibilities:
If the support of σ_2 includes b_2, then a_1 is dominated by a_3 and in that case, b_2 is dominated by b_1, a contradiction.
If the support of σ_2 contains only b_1 and b_3, then the value of the game is at least p. Denote α to be the probability that σ_2 assigns to b_1 and $1 - \alpha$ to be the probability it assigns to b_3. Then, it must be that firm 1's expected payoff from a_2 (which by Step 2 is in σ_1's support) is at least p, that is, $\alpha \cdot 1 + (1-\alpha)T_3 \geq p$, which implies that $\alpha \geq \frac{p-T_3}{1-T_3} = (1-p)$. Again using the inequality $(1-p)T_2 + pT_1 < p$, it follows that a_4 is not in the support of σ_1. However, in that case, b_3 is superior to b_1, a contradiction.

Step 4: *σ_1's support includes a_1.*
Otherwise, σ_2's support does not include b_2 (which is dominated by b_1), and in that case firm 1's statement a_1 obtains an expected payoff of more than p (since firm 2 also uses b_4), a contradiction to part (i).

Step 5: *σ_2's support includes b_2.*
Otherwise, the value of the game would be higher than p since by Step 4, a_1 is in σ_1's support and yields an expected payoff higher than p, since by Step 3, b_4 is in σ_2's support, a contradiction to part (i). ■