On Optimal Rules of Persuasion*

Jacob Glazer
Faculty of Management, Tel Aviv University
and
Department of Economics, Boston University
glazer@post.tau.ac.il

and

Ariel Rubinstein
School of Economics, Tel Aviv University
rariel@post.tau.ac.il

first version: 1 May 2003
revised version: 3 June 2003
final version: 1 November 2003

*Before reading the paper we advise you to play our "Persuasion Game" on-line : http://gametheory.tau.ac.il/exp5/

We are grateful to Noga Alon of Tel Aviv University for numerous discussions and assistance regarding Proposition 2. We also thank the co-editor of this journal, two anonymous referees, Andrew Caplean and Bart Lipman for valuable comments.

Abstract

A speaker wishes to persuade a listener to accept a certain request. The conditions under which the request is justified, from the listener's point of view, depend on the values of two aspects. The values of the aspects are known only to the speaker and the listener can check the value of at most one. A mechanism specifies a set of messages that the speaker can send and a rule which determines the listener's response, namely, which aspect he checks and whether he accepts or rejects the speaker's request. We study mechanisms that maximize the probability that the listener accepts the request when it is justified and rejects the request when it is unjustified, given that the speaker maximizes the probability that his request is accepted. We show that finding the optimal mechanism is equivalent to solving a linear programming problem in which the set of constraints is derived from what we call the L-principle.

Key words: persuasion, mechanism design, hard evidence, debates.

Classifications: C610, C72, D78, B201, R316.

1. **Introduction**

Our model deals with the situation in which one agent (*the speaker*) wishes to persuade another agent (*the listener*) to take a certain action. Whether or not the listener should take the action is dependent on information possessed by the speaker. The listener can obtain bits of relevant information but is restricted as to the total amount of evidence he can accumulate. The speaker can use only verbal statements to persuade. Whether or not the listener is persuaded is dependent on both the speaker's arguments and the hard evidence the listener has acquired.

Following are some real life examples:

A worker wishes to be hired by an employer for a certain position. The worker tells the employer about his previous experience in two similar jobs. The employer wishes to hire the worker if his average performance in the two previous jobs was above a certain minimal level. However, before making the final decision the employer has sufficient time to thoroughly interview at most one of the candidate's previous employers.

A suspect is arrested on the basis of testimonies provided by two witnesses. The suspect's lawyer claims that their testimonies to the police have serious inconsistencies and therefore his client should be released. The judge's preferred decision rule is to release the suspect only if the two testimonies substantially contradict one another; however, he is able to corroborate at most one of the two testimonies.

A doctor claims that he has correctly used two procedures to treat a patient who suffers from two chronic illnesses. An investigator on the case is asked to determine whether the combination of the two procedures was harmful. The investigator has access to the doctor's full report but verifying the details of more than one procedure would be too costly.

A decision maker asks a consultant for advice on whether or not to take on a particular project. The decision maker knows that the consultant is better informed about the state of the world than he is, but he also knows that it is in the consultant's interests that the project be carried out regardless of the state of the world. The decision maker is able to verify only a restricted number of facts that the consultant claims to be true.

In our model the listener has to choose between two actions *a* and *r*. A speaker's type is a realization of two aspects initially known only to the speaker. The listener's preferred action critically depends on the speaker's type whereas the speaker would like the listener to choose the action *a* regardless of his type.

We study a family of mechanisms in which the speaker sends a message to the listener and the listener can then choose to ascertain the realization of at most one of the two aspects. On the basis of the speaker's message and the acquired "hard" evidence, the listener is either persuaded to take the speaker's preferred action *a* or not. More specifically, a mechanism is composed of three elements: a set of messages the speaker can choose from; a function that specifies which aspect is to be checked depending on the speaker's message; and the action the listener finally takes as a function of the message sent by the speaker and the acquired information.

Two types of mechanisms will serve a special role in our analysis:

A) Deterministic mechanisms – for each of the two aspects certain criteria are determined and the speaker's preferred action is chosen if he can show that his type meets these pre specified criteria in at least one of the two aspects. In the first example above, a deterministic mechanism would be equivalent to asking the worker to provide a reference from one of his two previous employers which meets certain criteria.

B) Random mechanisms – the speaker is asked to report his type; one aspect is then chosen randomly and checked; and the action *a* is taken if and only if the speaker's report is not refuted. Returning to the first example above, a random mechanism would involve first asking the worker to justify his application by reporting his performance in each of his previous two jobs. Based on his report, the employer then randomly selects one of the two previous employers to interview and accepts the applicant if his report is not refuted.

We are interested in the properties of the mechanisms that are optimal from the point of view of the listener, namely, those in which it is least likely that the listener will choose the wrong action given that the speaker maximizes the probability that the action *a* will be taken. In our scenario, the listener does not have tools to deter the speaker from cheating and thus we can expect that the speaker will always argue that his information indicates that the action *a* should be taken. The problem therefore is to decide which rules the listener should follow in order to minimize the probability of making a mistake.

The main results of the paper are as follows:

1) Finding an optimal mechanism can be done by solving an auxiliary linear programming problem. The objective in the auxiliary problem is to minimize the probability of a mistake. The constraints are derived from a condition that we call the *L*-principle which can be demonstrated using the first example above: Assume that the worker's performances in each job is classified as good or bad and that the employer wishes to hire the worker only if his performance in both previous jobs was good. Consider the worker's three types: his performance was good in two previous jobs, good only in the first job and good only in the second job. The *L*-principle says that for any mechanism, the sum of the probabilities of a mistake conditional on each of the three worker's types is at least one.

2) An optimal mechanism with a very simple structure always exists. First, the speaker is asked to report his type. If the speaker admits that the action r should be taken, then the listener chooses r. If the speaker claims that the action a should be taken, the listener tosses a fair coin where on each of its two sides one of three symbols, r, 1 or 2 appears (the degenerate case where the same symbol appears on both sides is not excluded). The meaning of the symbol r is that the listener chooses the action r. The meaning of the symbol $i = 1,2$ is that the listener checks aspect i and takes the action a if and only if the speaker's claim regarding this aspect is confirmed.

3) The optimal mechanism is credible, that is, there exists an optimal strategy for the speaker which induces beliefs that make it optimal for the listener to follow the mechanism. Furthermore, the speaker's optimal strategy can be derived from the dual (auxiliary) linear programming problem.

4) For the case that all types are equally likely we identify certain "convexity" and "monotonicity" conditions under which there exists an optimal deterministic mechanism.

2. **The Model**

Let $\{1,,,,n\}$ be a set of random variables which we call *aspects*. Most of the analysis will be conducted for $n = 2$. The realization of aspect k is a member of a set X_k. A *problem* is (X,A,p), where $\emptyset \neq A \subset X = \times_{k=1,,n} X_k$ and p is a probability measure on X. We use the notation, p_x for the probability of type x, that is $p_x = p(\{x\})$. For the case that X is infinite we relate to p_x as a density function. For simplicity, we assume that $p_x > 0$ for all x. A problem is *finite* if the set X is finite. There are two agents: the *speaker* and the *listener*. A member of X is called a (speaker's) *type* and is interpreted as a possible characterization of the speaker. The listener has to take one of two actions: a (accept) or r (reject). The listener is interested in taking the action a if the speaker's type is in A and the action r if the type is in $R = X - A$. The speaker, regardless of his type, prefers the listener to take the action a. The speaker knows his type while the listener only knows its distribution. The listener can *check*, that is, find out the realization of, at most one of the n aspects.

A *mechanism* is (M,f), where M is a set (of *messages*) and $f : M \to Q$ where Q is the set of all lotteries $< \pi_0, d_0; \pi_1, d_1; \ldots; \pi_n, d_n >$ where $(\pi_i)_{i=0,1,..,n}$ is a probability vector and $d_k : X_k \to \{a,r\}$ where $X_0 = \{e\}$ is an arbitrary singleton set (that is d_0 is a constant). An element in Q is interpreted as a possible response of the listener to a message. With probability π_0 no aspect is checked and the action $d_0 \in \{a,r\}$ is taken and with probability π_k $(k = 1,..,n)$ aspect k is checked and if its realization is x_k the action $d_k(x_k)$ is taken. Our choice of the set Q captures the assumptions that the listener can check at most one aspect and that the aspect to be checked can be selected randomly.

A *direct mechanism* is one where $M = X$. For a direct mechanism (X,f) we say that following a message m the *mechanism verifies aspect k with probability* π_k when $f(m) =< \pi_0, d_0; \pi_1, d_1; \ldots; \pi_n, d_n >$ is such that $d_k(x_k) = a$ iff $x_k = m_k$. The *fair random mechanism* is the direct mechanism according to which, for every $m \in A$, the speaker verifies each aspect with probability $1/n$ and, for every $m \in R$, he chooses the action r. A mechanism is *deterministic* if for every $m \in M$ the lottery $f(m)$ is degenerate (that is, for some k, $\pi_k = 1$).

For every lottery $q =< \pi_0, d_0; \pi_1, d_1; \ldots; \pi_n, d_n >$ and every type x define $q(x)$ to be the probability that the action a is taken when the lottery q is applied to type x, that is, $q(x) = \sum_{\{k|d_k(x_k)=a\}} \pi_k$. We assume that given a mechanism (M,f) a speaker of type x will choose a message that maximizes the probability that the action a is taken, namely he chooses a message $m \in M$ which maximizes $f(m)(x)$. Let μ_x be the probability that the listener takes the wrong action with respect to type x, assuming the speaker's behavior. That is, for $x \in R$ we have $\mu_x = \max_{m \in M} f(m)(x)$ and for $x \in A$ we have $\mu_x = 1 - \max_{m \in M} f(m)(x)$. Note that all solutions to the speaker's maximization problem induce the same probability of a mistake. We will refer to $(\mu_x)_{x \in X}$ as the vector of mistakes induced by the mechanism. The *mistake probability* induced by the mechanism is $\int_{x \in X} p_x \mu_x$.

Given a problem (X, A, p), an *optimal mechanism* is one that minimizes the mistake probability. Note that the mechanisms are evaluated according to the listener's interests while ignoring those of the speaker. It should be mentioned that we do not restrict the discussion to direct mechanisms and do not apply the revelation principle.

Following is a concrete example:

Example 1: Let $X_1 = X_2 = [0, 1]$, let $A = \{(x_1, x_2) | x_1 + x_2 \geq 1\}$ and let p be the uniform distribution.

If the listener chooses to ignore the speaker's message, the lowest probability of a mistake he can obtain is 1/4. This mistake probability can be achieved by a mechanism in which aspect 1 is checked with probability 1 and action a is taken iff aspect 1's value is at least 1/2 (formally, $M = \{e\}$ and $f(e)$ is the degenerate lottery where $\pi_1 = 1$ and $d_1(x_1) = a$ iff $x_1 \geq 1/2$).

In this example, letting the speaker talk can improve matters. Consider the following deterministic direct mechanism ($M = X$) characterized by two numbers z_1 and z_2. Following the receipt of a message (m_1, m_2), the speaker verifies the value of aspect 1 if $m_1 \geq z_1$ and verifies the value of aspect 2 if $m_1 < z_1$ but $m_2 \geq z_2$. If $m_k < z_k$ for both k the action r is taken. One interpretation of this mechanism is that in order to persuade the listener, the speaker has to show that the realization of at least one of the aspects is above some threshold (which may be different for each aspect). The set of types for which the listener's action will be wrong consists of the three shaded triangles shown in figure 1a:

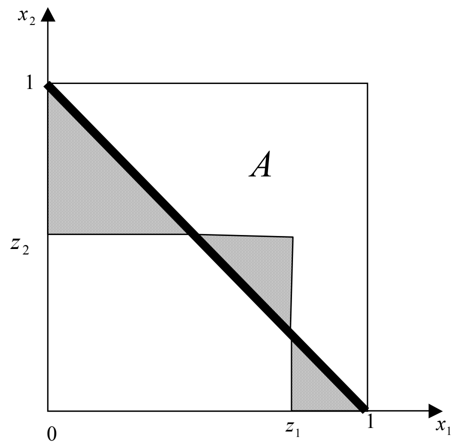

Figure 1a

One can see that the optimal thresholds are $z_1 = z_2 = 2/3$ yielding a mistake probability of 1/6. Is it possible to obtain a lower probability of a mistake by applying a non-deterministic mechanism? (Notice that the fair random mechanism does not help here as it yields mistake probability of 1/4.) We will return to this question later.

3. **A Basic Proposition**

From now on, assume $n = 2$. For simplicity of notation, we write μ_{ij} for $\mu_{(i,j)}$. The following proposition is key to our analysis (note that it is valid for both finite and infinite X and for any p):

Proposition 0: (The L principle) Let (X,A,p) be a problem. For any mechanism and for any three types $(i,j) \in A$, $(i,s) \in R$ and $(t,j) \in R$ it must be that $\mu_{ij} + \mu_{is} + \mu_{tj} \geq 1$.

Proof: Let (M,f) be a mechanism. Let m be a message optimal for type (i,j) and let $f(m) =< \pi_0, d_0; \pi_1, d_1; \pi_2, d_2 >$. For a proposition e, let δ_e be 1 if e is true and 0 if e is false. Then

$\mu_{ij} = \pi_0 \delta_{d_0=r} + \pi_1 \delta_{d_1(i)=r} + \pi_2 \delta_{d_2(j)=r}$.

If type (i,s) sends the message m (“claims” that he is (i,j)) the action a will be taken with probability $\pi_0\delta_{d_0=a} + \pi_1\delta_{d_1(i)=a}$ and therefore

$\mu_{is} \geq \pi_0\delta_{d_0=a} + \pi_1\delta_{d_1(i)=a}$.

Similarly,

$\mu_{tj} \geq \pi_0\delta_{d_0=a} + \pi_2\delta_{d_2(j)=a}$.

Therefore

$$\mu_{ij} + \mu_{is} + \mu_{tj} \geq \pi_0\delta_{d_0=r} + \pi_1\delta_{d_1(i)=r} + \pi_2\delta_{d_2(j)=r} + \pi_0\delta_{d_0=a} + \pi_1\delta_{d_1(i)=a} + \pi_0\delta_{d_0=a} + \pi_2\delta_{d_2(j)=a} = 1 + \pi_0\delta_{d_0=a} \geq 1.$$

■

The idea of the proof is as follows: whatever is the outcome of the randomization following a message m sent by type $(i,j) \in A$, either the mistaken action r is taken, or at least one of the two types (i,s) and (t,j) in R can induce the wrong action a by sending m.

We define an L to be any set of three types $(i,j) \in A$, $(i,s) \in R$ and $(t,j) \in R$. We refer to the result of Proposition 0 (the sum of mistakes in every L is at least 1) as *the L-principle*. Extending the L-principle to the case of $n > 2$ is done by defining an L to be a set of three types $x \in A$, $y \in R$ and $z \in R$ such that y and z each differ from x in the value of exactly one aspect.

4. **Examples**

In all our examples we take p to be uniform. For the case that X is finite we will refer to $\sum_{x \in X} \mu_x$ as the *number of mistakes*. When p is uniform and X is finite an optimal mechanism can be found by using a technique which relies on the L-principle: finding a mechanism that induces H mistakes and finding H disjoint L's allows us to conclude that this mechanism is optimal, thus yielding a mistake probability of $H/|X|$. The examples also give some intuition as to when optimality can be obtained by deterministic mechanisms and when it requires that the listener use randomization to determine the aspect to be checked.

Example 2

Let $X_1 = X_2 = \{1,..,5\}$ and $A = \{x | x_1 + x_2 \geq 7\}$. In Figure 2, each entry stands for an element in X and the types in A are indicated by the letter A. We mark 5 *disjoint* L's (the three elements of each L are indicated by the same number):

1	$A1$	A	A	A
2	5	$A2$	$A5$	A
3			$A3$	A
4			5	$A4$
	1	2	3	4

Figure 2

Following is a direct mechanism that induces 5 mistakes and is thus optimal: For any message m such that $m_k \leq 4$ for both k, the action r is taken. Otherwise, an aspect k, for which $m_k = 5$, is verified. In fact, this mechanism amounts to simply asking the speaker to present an aspect with a value of 5. The five mistakes are with respect to the three types $(3,4)$, $(4,4)$ and $(4,3)$ in A and the two types $(1,5)$ and $(5,1)$ in R.

Example 2 will later be generalized: when p is uniform constructing an optimal mechanism does not require randomization when the speaker's aim is to persuade the listener that the average of the values of the two aspects is above a certain threshold.

Example 3

This example shows that the conclusion of example 2 (randomization is not needed for the case in which the speaker tries to persuade the listener that the average of the values of the two aspects is above a certain threshold) does not hold for the case in which the number of aspects is greater than 2.

Consider the problem where $n = 3$, $X_k = \{0,1\}$ for $k = 1,2,3$ and $A = \{(x_1,x_2,x_3) | \sum_k x_k \geq 2\}$. Consider the mechanism in which the speaker is asked to name two aspects. The listener checks each of them with probability 1/2 and takes the action a if the value of the checked aspect is 1. This mechanism yields 1.5 mistakes (mistake probability of 3/16) since only the three types $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$ can each mislead the listener with probability 1/2.

To see that this is an optimal mechanism note that the following three inequalities hold:

$\mu_{(1,1,0)} + \mu_{(1,0,0)} + \mu_{(0,1,0)} \geq 1$

$\mu_{(1,0,1)} + \mu_{(1,0,0)} + \mu_{(0,0,1)} \geq 1$

$\mu_{(0,1,1)} + \mu_{(0,1,0)} + \mu_{(0,0,1)} \geq 1$

The minimum of $\sum_{x \in X} \mu_x$ subject to the constraint $\sum_{\{x|x_1+x_2+x_3=2\}} \mu_x + 2\sum_{\{x|x_1+x_2+x_3=1\}} \mu_x \geq 3$, implied by summing up the three inequalities, is attained when $\mu_x = 1/2$ for any $x \in \{(1,0,0),(0,1,0),(0,0,1)\}$ and $\mu_x = 0$ for any other $x \in X$. Thus, the number of mistakes cannot fall below 1.5.

Within the set of deterministic mechanisms an optimal mechanism involves taking the action a iff the speaker can show that either aspect 1 or aspect 2 has the value 1. This mechanism induces two mistakes with regard to types $(1,0,0)$ and $(0,1,0)$.

Example 4

This example corresponds to the situation described in the introduction in which a suspect's lawyer claims that the two testimonies brought against his client are inconsistent but the judge has time to thoroughly investigate only one of them. Consider the problem where $X_1 = X_2 = \{1,2,3,\ldots,I\}$ and $A = \{(x_1,x_2)\,|\,x_1 \neq x_2\}$. Intuitively, the knowledge of the value of only one aspect by itself is not useful to the listener. The optimal mechanism will be shown to be non-deterministic in this case.

The fair random mechanism (following a message in A each of the two aspects is verified with probability 0.5) induces $I/2$ mistakes. It is easy to show inductively that the minimal number of mistakes is $I/2$ starting from the two cases $I = 2$ and $I = 3$ illustrated in Figures 3a and 3b.

1	$R1$
$R1$	

Figure 3a

$*$	$*$	$R*$
$*$	$R*$	
$R*$		

Figure 3b

In the case of $I = 2$ there is one L. For $I = 3$, the maximal number of disjoint L's is one;

however, notice the six starred types - three in A and three in R. Each of the starred types in A combined with two of the starred types in R constitutes an L. Therefore, any mechanism induces mistake probabilities $(\mu_x)_{x\in X}$ satisfying:

$$\mu_{1,3} + \mu_{1,1} + \mu_{3,3} \geq 1$$

$$\mu_{1,2} + \mu_{1,1} + \mu_{2,2} \geq 1$$

$$\mu_{2,3} + \mu_{2,2} + \mu_{3,3} \geq 1$$

which imply that the sum of mistakes with respect to these six elements must be at least 1.5.

Any deterministic mechanism induces a vector $(\mu_x)_{x\in X}$ of mistakes with $\mu_x \in \{0,1\}$ for all x. If there is an i for which $\mu_{ii} = 0$, then for any $j \neq i$ either $\mu_{jj} = 1$ or $\mu_{i,j} = 1$ since if $\mu_{jj} = 0$ the constraint $\mu_{i,j} + \mu_{ii} + \mu_{j,j} \geq 1$ implies $\mu_{i,j} = 1$. Therefore $\sum_{x\in X} \mu_x \geq I - 1$. Thus, any deterministic mechanism induces at least $I - 1$ mistakes.

A deterministic mechanism that induces $I - 1$ mistakes is the one in which the speaker is asked to present an aspect whose realization is not 1 (thus yielding mistakes only for the types (i,i) with $i \neq 1$).

Example 5

Whereas in Example 4 the speaker tries to persuade the listener that the two aspects have different values, here he tries to persuade him that they have the same value, that is, $X_1 = X_2 = \{1,\ldots,I\}$ and $A = \{x|(x_1,x_2)\,|\,x_1 = x_2\}$. Here, it is also true that any information about one of the aspects provides no useful information to the listener but unlike the previous case, for $I > 2$, randomization is not helpful. The mechanism according to which the listener chooses r independently of the speaker's message without checking any of the aspects induces I mistakes. To see that one cannot reduce the number of mistakes note that the I sets $\{(i,i),$ $(i+1,i),(i,i+1)\}$ for $i = 1,..,I-1$ and $\{(I,I),(1,I),(I,1)\}$ consist of a collection of disjoint L's.

Comment: In example 5, with $I = 3$, the probability of a mistake is 1/3. This is in fact the worst case for the listener, that is, the optimal mechanism in our model with two aspects and any probability measure will never induce a mistake probability that is higher than 1/3. In fact, in every problem, both the "reject all" mechanism and the fair random mechanism guarantee a mistake probability of at most 1/3. If the probability of the set A is δ then the "reject all" mechanism yields mistake probabilities of δ, the fair random mechanism yields the mistake probability $(1-\delta)/2$ and $\min\{\delta, (1-\delta)/2\} \leq 1/3$.

Example 6

As mentioned above, the optimality criterion we employ involves maximizing the probability that the listener makes the right decision from his point of view while ignoring the interests of the speaker. If the optimality criterion also took into account the speaker's interests, the optimal mechanism would of course change. In particular, as this example shows, the listener might be indifferent between two optimal mechanisms while the speaker might not.

Consider the problem with $X = \{1,..,5\} \times \{1,..,5\}$ and $A = \{(x_1,x_2)|x_1+x_2 \in \{6,8,10\}\}$. The minimal number of mistakes is easily shown to be 8 and is obtained by both the deterministic mechanism "show me an aspect whose value is 5" and the fair random mechanism. However, under the fair random mechanism the listener can induce the action a with probability 17/25 whereas under the optimal deterministic mechanism he can do so only with probability 9/25. Thus, the fair random mechanism is superior for the speaker.

Let us return to example 1 and demonstrate the usefulness of the L-principle also in cases where the problem is not finite.

Example 1 (again): We have already found a deterministic mechanism with mistake probability of 1/6. To see that the mistake probability of any mechanism is at least 1/6, divide the unit square into 9 equal squares and divide each square into two triangles as shown in Figure 1b.

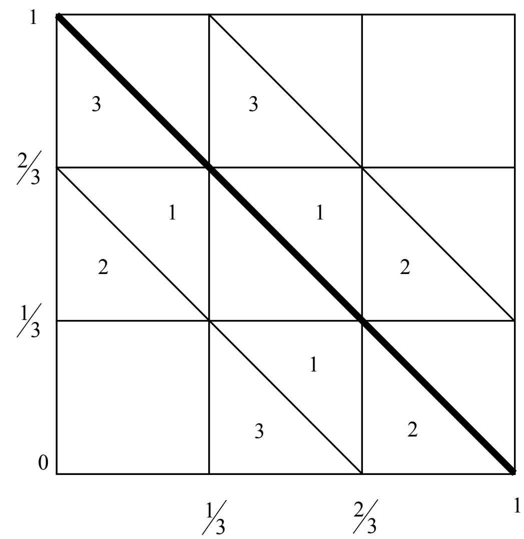

Figure 1b

The set $T_1 = \{(x_1,x_2) \in A | x_1 \leq 2/3 \text{ and } x_2 \leq 2/3\}$ is one of the three triangles denoted in the figure by the number 1. Any three points $x = (x_1,x_2) \in T_1$, $y = (x_1 - 1/3, x_2) \in R$ and $z = (x_1, x_2 - 1/3) \in R$ establish an L. By Proposition 0, $\mu_x + \mu_y + \mu_z \geq 1$. The collection of all these L's is a set of disjoint sets whose union is the three triangles denoted in the figure by the number 1. Therefore the integral of μ over these three triangles must be at least the size of T_1, namely 1/18. Similar considerations regarding the three triangles denoted by the number 2 and the three triangles denoted by the number 3 imply that $\int_{x \in X} \mu_x \geq 1/6$.

5. An Equivalent Linear Programming Problem

We will now show that for a finite problem, finding an optimal mechanism is equivalent to solving an auxiliary linear programming problem.

Let (X,A,p) be a finite problem. Define $P(X,A,p)$ to be the linear programming problem:

$$\min \sum_{x \in X} p_x \kappa_x$$

s.t. $\kappa_{ij} + \kappa_{is} + \kappa_{tj} \geq 1$ for all $(i,j) \in A$, $(i,s) \in R$ and $(t,j) \in R$

and $0 \leq \kappa_x$ for all $x \in X$.

We will show that the solution to $P(X,A,p)$ coincides with the vector of mistake probabilities induced by an optimal mechanism.

Note that not every vector which satisfies the constraints of $P(X,A,p)$, even if we add the constraints $\kappa_x \leq 1$ for all x, can be induced by a mechanism. "Incentive compatibility" implies additional constraints on the vector of mistake probabilities, $(\mu_x)_{x \in X}$, induced by a mechanism. For example, if $(i,j) \in A$, $(i,s) \in A$ and $(t,j) \in A$ then it is impossible that $\mu_{ij} = 0$ while both $\mu_{is} = 1$ and $\mu_{tj} = 1$, since at least one of the types, (i,s) or (t,j), can increase the probability that the action taken is a by imitating type (i,j). Nevertheless, we will show that any solution to the linear programming problem can be induced by a mechanism.

Proposition 1: Let (X,A,p) be a finite problem and let $(\kappa_x)_{x \in X}$ be a solution to $P(X,A,p)$. Then, there is an optimal mechanism such that the vector of mistakes induced by the mechanism is $(\kappa_x)_{x \in X}$.

Proof:

Step 1: Note that for every $x \in X$ it must be that $\kappa_x \leq 1$ and either $\kappa_x = 0$ or there are two other types y and z such that $\{x,y,z\}$ establish an L and $\kappa_x + \kappa_y + \kappa_z = 1$. Otherwise we could reduce κ_x and stay within the constraints.

Step 2: By Proposition 0 any vector of mistake probabilities induced by a mechanism satisfies the constraints. Thus, it is sufficient to construct a mechanism such that $(\mu_x)_{x \in X}$, the vector of its induced mistake probabilities, is equal to $(\kappa_x)_{x \in X}$.

Choose $M = X$. We use the convention $min_{x \in \emptyset} \kappa_x = 1$. For any message in R the action r is chosen.

For a message $(i,j) \in A$, distinguish between two cases:

(i) $\kappa_{ij} > 0$

-with probability κ_{ij} the action r is taken.

-with probability $min_{\{s|is\in R\}}\kappa_{is}$ the first aspect is verified.

-with probability $min_{\{t|tj\in R\}}\kappa_{tj}$ the second aspect is verified.

By step 1, $\kappa_{ij} + min_{\{s|is\in R\}}\kappa_{is} + min_{\{t|tj\in R\}}\kappa_{tj} = 1$.

(ii) $\kappa_{ij} = 0$

Note that $min_{\{s|is\in R\}}\kappa_{is} + min_{\{t|tj\in R\}}\kappa_{tj} \geq 1$. Choose two numbers $\alpha_1 \leq min_{\{s|is\in R\}}\kappa_{is}$ and $\alpha_2 \leq min_{\{t|tj\in R\}}\kappa_{tj}$ satisfying, $\alpha_1 + \alpha_2 = 1$. Aspect 1 (2) is verified with probability α_1 (α_2).

Step 3: We will now show that for the mechanism constructed in Step 2, $\mu_x \leq \kappa_x$ for every $x \in X$.

A type $(i,j) \in R$ cannot induce the action a with positive probability unless he sends a message $(i,s^*) \in A$ or $(t^*,j) \in A$. If he announces (i,s^*) the first aspect is verified with probability of at most $min_{\{s|is\in R\}}\kappa_{is} \leq \kappa_{ij}$. If he sends the message (t^*,j) the second aspect is verified with probability of at most $min_{\{t|tj\in R\}}\kappa_{tj} \leq \kappa_{ij}$. Thus, (i,j) cannot induce the action a with probability higher than κ_{ij}.

A type $(i,j) \in A$ who announces (i,j) will induce the action a with probability $1 - \kappa_{ij}$. Since his aim is to reduce the probability of a mistake, he will not induce a probability of a mistake higher than κ_{ij}.

Step 4 By Proposition 0, the vector $(\mu_x)_{x\in X}$ satisfies the constraints of $P(X,A,p)$ and by Step 3, the objective function assigns to this vector a value of at most $\sum_{x\in X} p_x\kappa_x$, which is the value of the solution to $P(X,A,p)$. Therefore it must be that $\mu_x = \kappa_x$ for all $x \in X$ and the mechanism we constructed in Step 2 is optimal. ■

6. **The Simple Structure of Optimal Mechanisms**

Next we will show that there always exists an optimal mechanism with a simple structure. In Step 2 of Proposition 1 we constructed a direct optimal mechanism in which the listener only verifies aspects. Thus, the fact that we allow the listener to condition his action on the exact value of the aspect he has checked does not enable him to reduce the mistake probability beyond what he could obtain were he only able to *verify* one aspect of the speaker's claim about his type. Proposition 1 by itself does not tell us anything about the probabilities used in an optimal mechanism. We will now show that for $n = 2$ one can always construct an optimal mechanism using only a fair coin as a form of randomization.

Proposition 2: For every finite problem (X, A, p) there exists an optimal mechanism which is direct (i.e. $M = X$) such that:

(a) If $m \in R$ the listener takes the action r whereas if $m \in A$ the listener does one of the following:

(i) takes the action r.

(ii) takes the action r with probability 1/2 and verifies one aspect with probability 1/2.

(iii) verifies each aspect with probability 1/2.

(iv) verifies one aspect with probability 1.

(Using our notation, for every $m \in R$, $f(m)$ is the degenerate lottery r and for every message $m = (m_1, m_2) \in A$, $f(m)$ is a lottery $< \pi_0, d_0; \pi_1, d_1; \pi_2, d_2 >$ where all π_i are in $\{0, 1/2, 1\}$, $d_0 = r$ and $d_i(x_i) = a$ iff $x_i = m_i$.)

(b) It is optimal for type $x \in A$ to report x and for type $x \in R$ to send a message y such that for one aspect k, $x_k = y_k$.

Proof: (a) A proposition due to Noga Alon (see Alon (2003)) states that if $(\alpha_x)_{x \in X}$ is an extreme point of the set of all vectors satisfying the constraints in $P(X, A, p)$, then $\alpha_x \in \{0, 1/2, 1\}$ for all $x \in X$. (Actually, our initial conjecture was that $\alpha_x \in \{0, 1\}$ for all $x \in A$ and $\alpha_x \in \{0, 1/2, 1\}$ for all $x \in R$. Noga Alon showed that we were only partially right and proved the modification of our conjecture).

Let $(\kappa_x)_{x \in X}$ be a solution to $P(X, A, p)$. As a solution to a linear programing problem the vector $(\kappa_x)_{x \in X}$ is an extreme point and thus $\kappa_x \in \{0, 1/2, 1\}$ for all $x \in X$. The construction of an optimal mechanism in Proposition 1 implies the rest of our claim since for every $i \in X_1$ and $j \in X_2$ the numbers $min_{\{s|is \in R\}} \kappa_{is}$ and $min_{\{t|tj \in R\}} \kappa_{tj}$ are all within $\{0, 1/2, 1\}$.

(b) The claim is straightforward for $x \in R$ and for $x \in A$ for which $\kappa_x = 0$. Type $(i,j) \in A$ for whom $\kappa_{ij} > 0$ can possibly obtain a positive probability of acceptance only by "cheating" about at most one aspect. If he claims to be type (t,j) then the probability that the second aspect will be verified is at most $min_{\{t|tj \in R\}} \kappa_{tj}$ which is exactly the probability that aspect 2 is verified when the speaker admits he is (i,j). ■

Example 7

In all previous examples the mistake probability of a type in A induced by the optimal mechanisms we have constructed was either 0 or 1. In this example any optimal mechanism induces a mistake probability of 0.5 for at least one type in A.

Let $X_1 = \{1,\ldots,8\}, X_2 = \{1,\ldots,7\}$. Types in A are denoted by A and p is uniform.

*	*	*	*	*	A *		
*	*	*	*	*	A *		
*	A	A	A	A	*	*	*
A	*	A	A	A	*	*	*
A	A	*	A	A	*	*	*
A	A	A	*	A	*	*	*
A	A	A	A	*	*	*	*

Figure 4a

9	10	1	2	5	15A	15	
8	6	3	4	7	16A	16	
11	A	1A	2A	14A	14	1	2
11A	11	3A	4A	A	15	3	4
A	6A	13	12A	5A	12	5	6
8A	A	13A	12	7A	13	7	8
9A	10A	A	A	14	16	9	10

Figure 4b

Any mechanism for this problem induces at least 16 mistakes since we can find 16 disjoint L's (see Figure 4b where each L is indicated by a distinct number). This number is at most 16 since the vector $\kappa_x = 1/2$ for any of the 32 types indicated by a star in Figure 4a and $\kappa_x = 0$ otherwise, satisfies the constraints of $P(X,A,p)$. Note that $\kappa_{66} = \kappa_{67} = 1/2$ although $(6,6)$ and $(6,7)$ are in A.

We will now show that for any $(\kappa_x)_{x \in X}$, a solution for $P(X,A,p)$, either κ_{66} or κ_{67} is not an integer.

First, note that $\kappa_{66} + \kappa_{67} \leq 1$. In Figure 4c we indicate 15 disjoint L's that do not contain any of the elements in the box $\{6,7,8\} \times \{6,7\}$ and thus the sum of mistakes in that box cannot exceed 1.

The 16 L's in Figure 4b do not contain $(8,6)$ and $(8,7)$ and thus $\kappa_{86} = \kappa_{87} = 0$. Similarly, $\kappa_{76} = \kappa_{77} = 0$.

9	10	1	2	5	A		
8	6	3	4	7	A		
13	$11A$	$1A$	$2A$	A	11	1	2
A	11	$3A$	$4A$	$12A$	12	3	4
$13A$	$6A$	14	A	$5A$	13	5	6
$8A$	A	$14A$	15	$7A$	14	7	8
$9A$	$10A$	A	$15A$	12	15	9	10

Figure 4c

Now assume that both κ_{66} and κ_{67} are integers. Then there is $j \in \{6,7\}$ so that $\kappa_{6j} = 0$. For any $i = 1,..,5$ it must be $\kappa_{6j} + \kappa_{6i} + \kappa_{7j} \geq 1$ and thus $\kappa_{6,i} = 1$. However, none of the 12 disjoint L's in Figure 4d contain any of the 5 types $(6,i)$ where $i = 1,..,5$ and hence the total number of mistakes is at least 17 which is a contradiction !

9	10	1	2	5	A		
8	6	3	4	7	A		
11	$11A$	$1A$	$2A$	A		1	2
A	11	$3A$	$4A$	A		3	4
A	$6A$	12	$12A$	$5A$		5	6
$8A$	A	A	12	$7A$		7	8
$9A$	$10A$	A	A			9	10

Figure 4d

7. **The Listener's Credibility and the Dual Problem**

In the construction of the optimal mechanism we have assumed that the listener is committed to the mechanism. It is possible, however, that the listener will calculate the optimal strategy of the speaker given the mechanism and will make an inference from the speaker's message about his type which will lead the listener to prefer not to follow the mechanism.

In other words, one can think about the situation as an extensive game: the speaker first sends a message to the listener. Having received the message the listener then *chooses* which aspect to check and once he has observed the realization of the aspect he *decides* whether to take the action *a* or *r*. A mechanism can be thought of a listener's strategy in this extensive game. One may ask whether the listener's strategy, which corresponds to an optimal mechanism, is part of a sequential equilibrium for this extensive game. If it is, we will say that the mechanism is *credible*.

We do not think that the sequential optimality of the listener's mechanism is a crucial criterion for its plausibility. The listener's commitment to the mechanism may arise from considerations external to the model (such as the desire to maintain his reputation). Note also that in our model sequential equilibrium does not impose any restrictions on the beliefs following messages outside the support of the speaker's strategy. This fact makes sequential rationality a rather weak restriction on the listener's strategy.

Nevertheless, the study of the "sequential rationality" of the listener's mechanism yields a surprising result. As we will now see, a solution to the dual linear programing problem of the primal problem studied in Proposition 1 can be transformed into a strategy for the speaker that, together with the listener's strategy as derived in Proposition 1, yields a sequential equilibrium.

The Dual Problem: Let (X,A,p) be a problem and let $T(X,A)$ be the set of all its L's. The dual problem to $P(X,A)$ is $D(X,A,p)$:

$\max \sum_{\Delta \in T(X,A)} \lambda_\Delta$

s.t. $\sum_{\{\Delta \in T(X,A) | x \in \Delta\}} \lambda_\Delta \leq p_x$ for all $x \in X$

and $0 \leq \lambda_\Delta$ for all $\Delta \in T(X,A)$.

Recall the examples in Section 4 where $p_x = 1/|X|$ for all $x \in X$. In the analysis of some of these examples we found a number of disjoint L's equal to the number of mistakes induced by some mechanism. Finding a collection of disjoint L's is equivalent to finding a point within the constraints of $D(X,A,p)$ (for which $\lambda_\Delta = 1/|X|$ for any Δ in the collection and $\lambda_\Delta = 0$ for Δ not in the collection). Finding a vector of mistake probabilities induced by a mechanism is equivalent to finding a point within the constraints of $P(X,A,p)$. Thus, our technique is equivalent to the technique commonly used in solving a linear programming problem based on the fact that the values of the solutions of $P(X,A,p)$ and $D(X,A,p)$ coincide.

The analysis of the case in which $I = 3$ in Example 4 can also be viewed in these terms. Assigning $\lambda_\Delta = 1/18$ for the three L's $\{(1,3),(1,1),(3,3)\}$, $\{(1,2),(1,1),(2,2)\}$ and $\{(2,3),(2,2),(3,3)\}$ and $\lambda_\Delta = 0$ otherwise, we identify a point in the constraints of the dual problem with an objective function value of 1/6. The fair random mechanism induces the mistake probabilities of $\mu_x = 1/2$ for the three points on the main diagonal and $\mu_x = 0$ otherwise, yielding the value 1/6 for the objective function of the primal problem.

Proposition 3: Let (X,A,p) be a finite problem. An optimal mechanism, built in Propositions 1 and 2 from a solution $(\kappa_x)_{x\in X}$ to $P(X,A,p)$ satisfying that $(\kappa_x)_{x\in X} \in \{0,1/2,1\}$ for all $x \in X$, is credible.

Proof: By the Complementary Slackness Theorem the dual problem $D(X,A,p)$ has a solution $(\lambda_\Delta)_{\Delta\in T(X,A)}$ such that $\lambda_\Delta(1 - \sum_{x\in\Delta}\kappa_x) = 0$ for all $\Delta \in T(X,A)$ and $\kappa_x(p_x - \sum_{x\in\Delta\in T(X,A)}\lambda_\Delta) = 0$ for all $x \in X$. The solutions to the primal and dual problems have the same value, that is $\sum_{\Delta\in T(X,A)}\lambda_\Delta = \sum_{x\in X}p_x\kappa_x$.

Consider the following speaker's strategy: Let x be a speaker's type. Every $x \in A$ announces x with certainty. Every $x \in R$ announces x with probability $1 - \sum_{\{\Delta\in T(X,A)|x\in\Delta\}}\lambda_\Delta/p_x$ and every Δ for which $x \in \Delta$ contributes λ_Δ/p_x to the probability that he announces the type in $A \cap \Delta$. This strategy is well-defined since by the dual problem constraints $\sum_{\{\Delta\in T(X,A)|x\in\Delta\}}\lambda_\Delta \le p_x$.

We first show that this strategy is a speaker's best response to the listener's strategy. By Proposition 2 we know that it is optimal for every $x \in A$ to announce x. Let $x \in R$. If $\kappa_x = 0$, type x cannot induce the listener to choose a with a positive probability and any strategy for type x is thus optimal. If $\kappa_x > 0$, then $\sum_{\{\Delta\in T(X,A)|x\in\Delta\}}\lambda_\Delta = p_x$. If a message $z \in A$ is sent by x with a positive probability, then there exists an L, $\Delta = \{z \in A, x \in R, y \in R\}$ for which $\lambda_\Delta > 0$ and thus $\kappa_z + \kappa_x + \kappa_y = 1$. Following are three configurations to consider:

(1) $\kappa_z = 0$ $\kappa_x = \kappa_y = 1/2$

On receiving the message z the listener verifies each of the aspects with probability 1/2. Thus, by sending the message z, type x will induce a with probability 1/2 which is the best he can do.

(2) $\kappa_z = 1/2$ $\kappa_x = 1/2$ and $\kappa_y = 0$

The listener takes the action r with probability 1/2 and verifies with probability 1/2 the aspect k for which $z_k = x_k$. By announcing z, type x induces the action a with probability 1/2 which is the best he can do.

(3) $\kappa_z = 0$, $\kappa_x = 1$ and $\kappa_y = 0$

The listener verifies with certainty the aspect k for which $z_k = x_k$. Thus, by announcing z type x induces a with probability 1.

It remains to show that this strategy rationalizes the listener's strategy.

Assume the message $m \in R$ is sent. There is no case in which a message $m \in R$ is sent by a type in A and thus we can assign to the listener the belief that it was sent by a type in R and therefore choosing r is indeed optimal.

Assume $m \in A$ is sent. We distinguish between two cases:

(i) $\kappa_m > 0$. It must be that $\sum_{m \in \Delta \in T(X,A)} \lambda_\Delta = p_x$. The induced Bayesian beliefs assign probability 1/3 to each of the following three events: the speaker is $m \in A$, the speaker is a type in R which shares with m the value of the first aspect, and the speaker is a type in R which shares with m the value of the second aspect. Conditional on these beliefs, the listener is indifferent between verifying one of the aspects and choosing r, each of which induces a mistake probability of 1/3.

(ii) $\kappa_m = 0$. The induced Bayesian beliefs assign equal probabilities to the event that the speaker is a type $x \in R$ and $x_1 = m_1$ and to the event that the speaker is a type $x \in R$ and $x_2 = m_2$. This probability is not higher than the probability the listener assigns to the event that the speaker is of type m. Thus, verifying any one of the aspects is optimal.

8. **The Optimality of Deterministic Mechanisms**

One can think about a deterministic mechanism in the following way: once the speaker has sent a message m, the listener checks one aspect $k(m)$ with probability 1 and chooses a if and only if the value of the aspect is in some set $V(m) \subseteq X_{k(m)}$. A speaker of type (x_1,x_2) will be able to induce the listener to take the action a if and only if there is a message m such that $x_{k(m)} \in V(m)$. Denote $V_k = \cup_{k(m)=k} V(m)$. A type (x_1,x_2) will induce a if and only if $x_k \in V_k$ for at least one k. Thus, for any deterministic mechanism there are two sets $V_1 \subseteq X_1$ and $V_2 \subseteq X_2$ such that the probability of a mistake is the probability of $\{(x_1,x_2) \in A \mid \text{for no } k, x_k \in V_k\} \cup \{(x_1,x_2) \in R \mid \text{for at least one } k, x_k \in V_k\}$. We call V_1 and V_2 the *sets of persuasive facts*.

We now derive a simple necessary condition for a mechanism to be optimal within the set of deterministic mechanisms:

Proposition 4: Let (X,A,p) be a finite problem. For a mechanism to be optimal within the set of deterministic mechanisms, its sets of persuasive facts V_1 and V_2 must satisfy:

for any $x_1 \in V_1$ $p\{(x_1,x_2) \in A \mid x_2 \notin V_2\} \geq p\{(x_1,x_2) \in R \mid x_2 \notin V_2\}$ and

for any $x_1 \notin V_1$ $p\{(x_1,x_2) \in A \mid x_2 \notin V_2\} \leq p\{(x_1,x_2) \in R \mid x_2 \notin V_2\}$

Similar conditions hold for V_2.

Proof: Assume, for example, that $s \in V_1$ but that $p\{(s,x_2) \in A \mid x_2 \notin V_2\} < p\{(s,x_2) \in R \mid x_2 \notin V_2\}$. Eliminating s from V_1 will decrease the mistake probability. To see this, note first that every type x such that either $x_1 \neq s$ or $x_2 \in V_2$ can induce the action a iff he could induce it prior to the elimination of s from V_1. Any type x such that $x_1 = s$ and $x_2 \notin V_2$ can induce the action a prior to the elimination but cannot do so following it. Thus, elimination of such an s reduces the mistake probability. ■

The condition stated in Proposition 4 is necessary but not sufficient for a mechanism to be optimal within the set of deterministic mechanisms: Returning to example 4 with $X_1 = X_2 = \{1,2,3,4\}$, a mechanism with $V_1 = V_2 = \{3,4\}$ satisfies the conditions in the proposition and yields 4 mistakes, while the mechanism with $V_1 = V_2 = \{2,3,4\}$ yields only 3 mistakes.

Finally, for problems with uniform probability we will identify conditions that guarantee that there exists an optimal mechanism which is deterministic. Let $X \subseteq \Re^2$. We say that a set $A \subseteq X$ is *monotonic* if for every $s > s'$ and for every t, $(s',t) \in A$ implies $(s,t) \in A$ and

$(t,s') \in A$ implies $(t,s) \in A$. In other words, a set is monotonic if, for every aspect, the higher its value the better indication it is that the type is in A. The sets A in Examples 1 and 2 are monotonic whereas the sets A in examples 4, 5, 6 and 7 are not.

The following proposition refers to the case in which the set of types is a continuum although it also gives some insight into the finite case:

Proposition 5: Let $X = [0,1] \times [0,1]$ and p be uniform and assume that A is monotonic and that R is closed, convex and non-empty. Then there exists an optimal mechanism which is direct and deterministic with (sets of persuasive facts) $V_k = [z_k, 1]$ for some $z_k > 0$ for both k.

Proof: The mistake probability induced by the deterministic mechanisms with $V_k = [y_k, 1]$ is continuous in y_1 and y_2. Thus, there is an optimal mechanism within this class characterized by the sets of persuasive facts $V_k = [z_k, 1]$. By our assumptions about R it must be that $z_k > 0$ for both k.

Assume first that $z_k < 1$ for both k. It follows from a modification of Proposition 4 that $z_2/2 = \max\{s | (z_1, s) \in R\}$ and $z_1/2 = \max\{s | (s, z_2) \in R\}$.

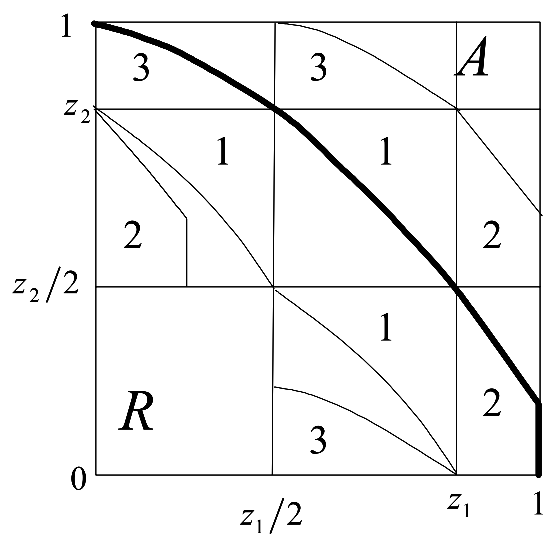

Figure 5

The rest of the proof extends the idea embedded in Example 1. In Figure 5, the set R is the area below the bold curve. The mistake probability induced by the mechanism above is the sum of the probabilities of the the following three disjoint sets:

$T_1 = \{(x_1,x_2) \in A | x_1 \leq z_1$ and $x_2 \leq z_2\}$

$T_2 = \{(x_1,x_2) \in R | x_1 > z_1$ and $x_2 \leq z_2/2\}$

$T_3 = \{(x_1,x_2) \in R | x_1 \leq z_1/2$ and $x_2 > z_2\}$.

Note that by the convexity of R the sets $\{(x_1 - z_1/2, x_2) | (x_1,x_2) \in T_1\}$ and $\{(x_1 - z_1, x_2 + z_2/2) | (x_1,x_2) \in T_2\}$ have an empty intersection. Similarly, $\{(x_1, x_2 - z_2/2) | (x_1,x_2) \in T_1\}$ and $\{(x_1 + z_1/2, x_2 - z_2) | (x_1,x_2) \in T_3\}$ are disjoint. The collection of all sets $\{x,(x_1 - z_1/2, x_2), (x_1, x_2 - z_1/2)\}$ with $x \in T_1$, $\{(x_1, x_2 + z_2/2), x, (x_1 - z_1, x_2 + z_2/2)\}$ with $x \in T_2$ and $\{(x_1 + z_1/2, x_2), x, (x_1 + z_1/2, x_2 - z_2)\}$ with $x \in T_3$ is a collection of disjoint L's. Thus, the mistake probability induced by any mechanism must be at least the probability of $T_1 \cup T_2 \cup T_3$.

As for the case where $z_1 = z_2 = 1$ it must be that $max\{s | (1,s) \in R\} \geq 1/2$ and $max\{s | (s,1) \in R\} \geq 1/2$ and thus the set A is a subset of $[1/2,1] \times [1/2,1]$. The mechanism "reject all" induces a mistake probability equal to the probability of the set A. Any mechanism induces a mistake probability at least as large as the probability of A since the collection of all sets $\{x,(x_1 - 1/2, x_2), (x_1, x_2 - 1/2)\}$ with $x \in A$, is a collection of disjoint L's.

Arguments similar to the above complete the proof for the case in which $z_k = 1$ for one k.

■

When the problem is monotonic but the set R is not convex, the optimal mechanism may not be deterministic. The following example is finite but could be extended to the case of a continuum.

Example 8

Let $X_1 = X_2 = \{1,2,..,5\}$. The elements in R are indicated in Figure 6.

$R*$	$*$			$*$
$R1$			1	
$R3$	3			
$R2$	$R*$	2		$*$
R	$R3$	$R2$	$R1$	$R*$

Figure 6

We will see that the optimal mechanism yields 4.5 mistakes and thus must be non-deterministic. Notice the six starred elements that produce three (non-disjoint) L's. Any mechanism which induces mistake probabilities $(\mu_x)_{x\in X}$ must satisfy:

$\mu_{5,5} + \mu_{1,5} + \mu_{5,1} \geq 1$

$\mu_{2,5} + \mu_{2,2} + \mu_{1,5} \geq 1$

$\mu_{5,2} + \mu_{2,2} + \mu_{5,1} \geq 1$

which imply that the sum of mistakes with respect to these six types must be at least 1.5. At least three additional mistakes must be induced with respect to the three disjoint L's indicated by the numbers 1, 2 and 3 in the figure. The fair random mechanism yields 4.5 mistakes (the 9 types in $R - \{(1,1)\}$ induce the action a with probability 0.5) and thus is optimal.

9. **Related Literature**

Our paper is related to the literature on strategic information transmission (see, for example, Crawford and Sobel (1982)). This literature studies a model in which a sender sends a costless message to a receiver. The listener cannot verify any of the information possessed by the sender. The interests of the sender and the receiver do not necessarily coincide. The situation is analyzed as a game and the main question asked is whether an informative sequential equilibrium exists. In contrast, the speaker in our model can choose to check some of the relevant information and the situation is analyzed as a mechanism design problem.

Some papers have studied principal agent problems in situations where the principal can obtain "hard evidence". In a very different context, Townsend (1979) studied the structure of efficient contracts in a model where the principal insures the agent against variability in the agent's wealth. The transfer of money from one party to the other may depend on the agent's wealth which is initially known only to the agent. The principal can verify the agent's wealth if he incurs some cost. In this model the choice of the principal is whether or not to verify the state whereas in our model the focus is on the principal's choice of which aspect to check.

Some of the literature has studied the circumstances under which the revelation principle holds, whereas our main interest is in characterizing the optimal mechanisms.

Green and Laffont (1986) studies mechanisms in which the set of messages each type can send depends on the type and is a subset of the set of types. Their framework does not allow the listener to randomize. Furthermore, their model does not cover the case in which the speaker can show the value of the realization of one of the aspects. In particular, assuming in their framework that a type (i,j) can only send messages like (i,s) or (t,j) is not the same as assuming that he can present one of the aspects. The reason is that a message (m_1, m_2) would not reveal whether the agent actually showed that the realization of aspect 1 is m_1 or that he showed that the realization of aspect 2 is m_2.

In Bull and Watson (2002) an agent can also show some evidence. A key condition in their paper is what they call "normality": if type x can distinguish himself from type x' and from x'' then he can also distinguish himself from both, a condition that does not hold in our framework. Furthermore, they do not consider randomized mechanisms.

A related paper is Fishman and Hagerty (1990). One interpretation of what they do is the analysis of the optimal deterministic mechanisms for the problem $(\{0,1\}^n, \{x|\sum_k x_k > b\})$ for some b.

This paper is rooted in Glazer and Rubinstein (2001) in which we study the design of optimal deterministic debate mechanisms in a specific example. (Other models of optimal design of debate rules with hard evidence are Shin (1994) and Lipman and Seppi (1995).) The two models are quite different but nevertheless have some common features. In both models there is a listener and speaker(s); the listener takes an action after listening to arguments made by the speaker(s); an instance is characterized by the realization of several aspects and the speaker(s) know(s) the realization of the aspects while the listener does not; a constrained amount of "hard" evidence can be revealed by the speaker(s) or checked by the listener; and finally, the listener must base his decision on only partial information. In both papers we look for a mechanism that minimizes the probability that the listener will take the wrong action.

References

Alon, Noga (2003) Problems and Results in Extremal Combinatorics - II, (memo).

Bull, Jesse and Joel Watson (2002), Hard Evidence and Mechanism Design, working paper.

Crawford, Vincent P. and Joel Sobel (1982), Strategic Information Transmission, *Econometrica*, 50, 1431-1451.

Fishman, Michael J. and Kathleen M. Hagerty (1990), The Optimal Amount of Discretion to Allow in Disclosures, *Quarterly Journal of Economics*, 105, 427-444.

Glazer, Jacob and Ariel Rubinstein (2001), Debates and Decisions, On a Rationale of Argumentation Rules, *Games and Economic Behavior*, 36 (2001), 158-173

Green, Jerry and Jean-Jacques Laffont (1986), Partially Verifiable Information and Mechanism Design, *Review of Economic Studies*, 447-456.

Lipman, Barton L. and Duane J.Seppi, (1995), Robust Inference in Communication Games with Partial Provanility, *Journal of Economic Theory, 66, 370-405.*

Shin, Hyun Song (1994), The Burden of proof in a Game of Persuasion, *Journal of Economic Theory*, 64, 253-264.

Townsend, Robert (1979), Optimal Contracts and Competitive Markets with Costly State Versification, *Journal of Economic Theory*, 21, 265-293.