Magical Implementation

Jacob Glazer

School of Management, Tel Aviv University
and Department of Economics, The University of Warwick

Ariel Rubinstein

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

November 3, 2024

Abstract: A principal would like to decide which of two parties deserves a prize. Each party privately observes the state of nature that determines which of them deserves the prize. The principal presents each party with a text that truthfully describes the conditions for deserving the prize and asks each of them what the state of nature is. The parties can cheat but the principal knows their choice procedure. The principal is able to "magically implement" his goal if he can come up with a pair of texts satisfying that in any dispute, he will recognize the cheater by applying the "honest-cheater asymmetry principle". According to this principle, the truth is with the party satisfying that if his statement is true, then the other party (using the given choice procedure) could have cheated and made the statement he is making, but not the other way around. Examples are presented to illustrate the concept.

KEYWORDS: Magical implementation

JEL classification: D82

We thank Shlomo Naeh (Department of Jewish Thought, The Hebrew University of Jerusalem) for alerting us to the principle of "lectio difficilior potior". We also thank Kfir Eliaz, Avner Shaked. Rani Spiegler, Kemal Yildiz and especially Áron Tóbiás, for their comments.

1. Introduction

Two invigilators, H and L, have witnessed a student receiving a whispered message from another student during an exam. The invigilators have not seen the questions on the exam but would be able to solve them. It is known that H is hostile to the student who received the message while L likes him. The exam includes multiple questions but only one refers to the variable α and reads as follows: "Solve the equation $\alpha+1=4$." The student answers the question correctly. Invigilator H claims that the whispered message was: "$\alpha=3$." This is a serious allegation and if correct, the student's exam will be disqualified. Invigilator L claims that the whispered message was: "Solve the equation $\alpha+1=4$ first." If he is right, then the student's answer genuinely reflects his knowledge of the material and there will not be any serious consequences. Who should be believed: H or L?

Although there is no definitive proof one way or the other, we would choose to believe L. The reasoning would be that if the message was "Solve the equation $\alpha+1=4$ first", then H could solve the equation himself and claim that the message was "$\alpha=3$". On the other hand, if the message was "$\alpha=3$" it is very unlikely that L (who, as mentioned, has not seen the exam questions) could guess that the equation to be solved is $\alpha+1=4$ rather than any other equation with the same solution. Hence, there is an asymmetry between the two conflicting claims which makes it possible to reasonably conclude that L's claim is the truthful one.

In the above episode, the asymmetry between the two claims arises naturally and is not engineered by someone seeking to uncover the truth. In other cases, one might consider designing a mechanism that creates some sort of asymmetry between a truth-teller and a cheater, which the principal would be able to exploit in order to identify the truth-teller with reasonable certainty. The design of such a mechanism is at the core of our analysis.

We consider situations of the following nature: Two parties claim a prize being offered by a principal. The principal thinks that only one of them truly deserves the prize, and his identity is determined *unequivocally* by facts known exclusively to the two parties (but not to the principal). The parties do not know the criteria which guide the principal and yet both parties insist that they deserve the prize.

The situations we have in mind are related to the biblical Judgement of Solomon, where two women claim to be the mother of the same baby and the king must decide who is telling the truth. In that story, unlike ours, the women know the circumstances under which King Solomon wishes to deliver the baby to each of them. There is also an asymmetry in the women's preferences with regard to the potential consequences of the ruling: the true mother likes the baby "more" than the fake mother in the sense that the true mother – whichever woman she is – is willing to pay more for the baby than the fake mother. This asymmetry allows Glazer and Ma (1989) (later generalized by Perry and Reny (1999)) to construct a game form with the feature that regardless of who the true mother is, the induced extensive game has a unique subgame perfect equilibrium with the outcome that the true mother gets the baby without making any payment.

We consider similar disputes without assuming any asymmetry in preferences or information and in which neither party is aware of the conditions under which he deserves the prize. The only distinction between the two parties is that one deserves the prize and can claim it by telling the truth about the state while the other can claim the prize only by telling a lie.

The model is enriched by (i) a *langauge* that the principal can use to compose a text to describe the circumstances under which a party deserves the prize according to the principal's judgement, and (ii) a *choice procedure* that determines the state a party announces, given the text he received and the realized state he observed. The choice procedure is common to the two parties and is known to the principal. Both the language and the choice procedure are situation-specific. For now, their details are vague; nonetheless, each situation we analyze below will have a formal specification.

We are interested in mechanisms that generate the correct outcome, based on the asymmetry between a truth-teller and a cheater. The approach adapted is not game-theoretic but rather rests on a novel concept we refer to as "*magical implementation*". A magical implementation mechanism consists of the following stages (after the parties have both been informed about the state):

Stage 1: The principal provides each party with a true and full description of the set of states in which the party deserves the prize. Being constrained by the language, there are numerous texts that could describe this set and the particular text presented to a party is chosen at the principal's discretion.

Stage 2: Each party is required to present a (true or false) factual statement (a state) to the principal and is informed that if his statement does not justify his claim for the prize, then he will not receive it.
Stage 3: The principal considers the statements, s_1 and s_2, made by party 1 and party 2, respectively, and makes his decision as follows:

- If both statements imply that the same party deserves the prize, then he awards it to that party.
- If both parties make a statement justifying their own claim to the prize, then the principal checks whether there is a party i such that if s_i is true, then the choice procedure might have enabled party j to make the statement s_j, whereas if s_j is true, the procedure could not have enabled party i to make the statement s_i. In this scenario, the prize is awarded to party i.
- In all other cases, neither party is awarded the prize.

In other words, the principal provides each party with an accurate description of the conditions under which he deserves the prize. The principal will grant the prize if the two parties agree on who deserves the prize, or if he can apply what we refer to as the "*honest-cheater asymmetry principle*": the truth is with the party satisfying that if his statement is true, then the other party (using the given choice procedure) could have cheated and made the statement he is making, but not the other way around.

Asymmetries between statements are often used in practice as a tool to decide which of two conflicting statements is true. For example, scholars of old manuscripts who have before them two versions of the same text, but only one of which can be genuine, use such asymmetries as a tool to decide which one is the original. A principle called "Lectio difficilior potior" ("the more difficult reading is the stronger") instructs scholars to choose the more unusual text.

Another such natural asymmetry involves word associations. As Michelbacher, Evert, and Schütze (2007) put it: "Human word associations are asymmetric or directed." For example, they note that 'soup' triggers 'tomato' much more often than 'tomato' triggers 'soup'; 'mango' triggers 'fruit' much more often than the reverse; a person with the name David Churchill brings Winston Churchill to mind but not the other way around, etc. Therefore if two parties disagree about which university a certain professor graduated from, where one claims it is MIT and the other that it is NIT, then we would tend to believe that the professor graduated from NIT.

In what follows, we formalize the concept of magical implementation and apply it in three examples.

2. The formal framework

Parties 1 and 2 are in a dispute over a single, indivisible prize. Let S be the set of all possible states of the world. The set S is partitioned into two disjoint subsets, W^1 and W^2, where W^i denotes the set of states in which party i should win the prize. In every state both parties are informed about the state but do not know the partition that determines who deserves the prize. We refer to the tuple $\langle S, W^1, W^2\rangle$ as an *implementation problem*.

A principal who is not informed about the state needs to rely on the parties in order to award the prize correctly. He constructs a pair of texts T^1 and T^2, where T^i is the text provided to party i. We interpret a text as a description of circumstances in which the party that receives the text deserves the prize. After receiving the text T^i, party i sends a message to the principal in the form of a state in S.

In choosing the texts, the principal uses a *language* $\mathfrak{L} = \langle \mathfrak{T}, Int\rangle$, where $\mathfrak{T}$ is a set of *texts* and Int is an interpretation function that assigns to each $T \in \mathfrak{T}$ a subset $Int(T)$ of states in which T is true. A different language is considered in each section below.

The parties apply a choice procedure C that attaches to each text T and state s a non-empty subset of states $C(T,s)$. Given that a party receives a text T and observes s the procedure of choice may lead him to announce any state in $C(T,s)$. When $C(T,s) = \{t\}$, we write $C(T,s) = t$. It will be always assumed that $C(T,s) = s$ whenever s satisfies the text T (that is, $s \in Int(T)$).

For any text T, the choice procedure C induces a binary relation $\to_T$ on S such that $s \to_T t$ is interpreted as: "if the true state is s and the text T is not satisfied by s, then the choice procedure may lead the party to claim t which does satisfy T." That is, $s \to_T t$ if:

(i) $s \notin Int(T)$)

(ii) $t \in C(T,s)$; and

(iii) $t \in Int(T)$.

We say that the pair of texts (T^1, T^2) *magically implements* $\langle S, W^1, W^2 \rangle$ if:
(1) $Int(T^i) = W^i$ for both i. That is, the principal provides each party with a *correct* description of the circumstances under which he deserves the prize. However, there may be multiple texts T for which $Int(T) = W^i$, and the principal seeks a pair of texts that will enable him to grant the prize to the deserving party.
(2) Given any state $s \in W^i$ and for any state $t \in S$ such that $s \to_{T^j} t$ (and thus $t \in W^j$), we have $t \not\to_{T^i} s$. That is, given any state, if the undeserving party cheats, then the principal can apply the *honest-cheater asymmetry principle* and infer which party is telling the truth.

We use the term *magical implementation* because magicians possess skills to discern subtle patterns and behavioral cues in human actions, which they are able to exploit in order to create the illusion of a miracle. The principal, in his role as magician, exploits his understanding of human imperfections in order to achieve his goal. Whereas a magician wishes to entertain his audience, the principal wishes to identify which of two rival parties is telling the truth.

Notice that if the parties understand the principal's inference method, then following the choice procedure is not a dominating strategy and outsmarting the choice procedure might be beneficial. When $s \in W^i$, we assume that party j will declare a state t such that $s \to_{T^j} t$ even if $t \not\to_{T^i} s$. But in that case, party j would do better if he finds a state r such that $r \to_{T^i} s$ and $s \not\to_{T^j} r$ and based on this lie persuades the principal that party i is the cheater.

Discussion: The notion of magical implementation is fundamentally different from the classical notion of Nash implementation. For one thing, the mechanism does not involve a game. The principal provides each party with a text that truthfully describes the circumstances in which he deserves the prize and commits that the party will not receive the prize if his claim does not meet the conditions described in the text. The principal does not inform the parties of what will happen if their claims contradict each other. The parties do not think strategically, i.e. they do not take into consideration the other party's actions. Each party acts as a "problem solver" *being aware* that he will certainly not get the prize if he does not solve the problem and *without being aware* that even if he successfully solves the problem he may not get the prize.

The honest-cheater asymmetry principle differs from the basic idea behind the standard Nash implementation mechanism a la Maskin (1999). There, in an environment of at least three agents, the mechanism accepts an appeal by an agent if and only if it is against his own interests, according to a consensus among the other agents about his preferences. Unlike our mechanism, it does not take into account the interests of the other agents to cheat, given the report of the appealing agent. Obviously, Nash implementation is not feasible in our environment. Magical implementation is based on the asymmetry created by the use of the choice procedure rather than exploiting of differences in interests.

Notice the fundamental difference between our approach and that of the literature on implementation with hard evidence (see, for example, Green and Laffont (1986), Lipman and Seppi (1995), Glazer and Rubinstein (2006) and Ben-Porath, Dekel and Lipman (2019)). In that literature, an informed agent is limited as to the lies he can tell about the state. These limits are given and are not affected by the mechanism designed by the principal. In contrast, the set of states in which an agent considers cheating in our case is determined endogenously by the texts presented by the principal and the choice procedure.

3. Setting a trap for the cheater

3.1 A motivating story

A village in DrSeussLand is populated by Yooks and Zooks. An unknown person was seen planting a bomb. The police arrest 8 suspects in the neighborhood: 3 Yooks (denoted as y_1, y_2, y_3) and 5 Zooks (denoted as z_1, z_2, z_3, z_4, z_5). It is certain that one of the suspects is the terrorist. There are two witnesses to the event – one of them a Yook and the other a Zook. The witnesses are able to recognize the terrorist. They do not know whether he or any other suspect is a Yook or a Zook because you can't distinguish Yooks and Zooks by their appearance or names. This does not prevent the witnesses from being convinced that the terrorist is from the other group. The witnesses are reluctant to identify the terrorist because he could belong to their own group, and such an action would be seen as a betrayal. Thus, there is no way to identify the terrorist by simply asking the witnesses who it is.

The authorities know which group each of the suspects belong to and are aware of the witnesses' reluctance to turn in a member of their own group. Nevertheless, they are determined to identify the terrorist. They come up with the following magical implementation idea. They construct two websites – one for the Yook witness and the other for the Zook witness. On the first page of each site, they insert the pictures and names of the 8 suspects.

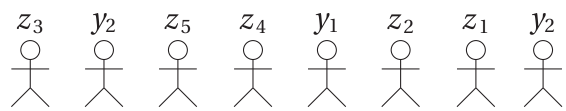

Figure 1. The first page of both websites

Each website consists of many pages, each of them displaying the names and pictures of two suspects. The left graph on Figure 2 describes the Yook's webiste while the right graph describes the Zook's. An edge in a graph means that the two connected suspects appear on one of the pages in the corresponding website. Both witnesses are obliged to name a suspect.

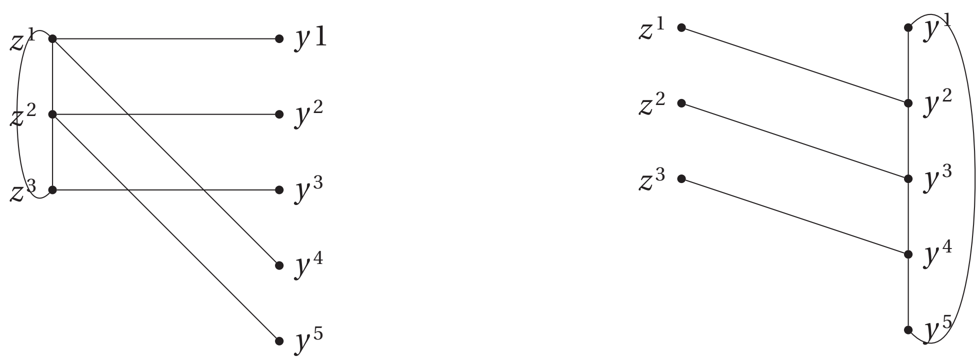

Figure 2. The graph on the left describes the Yook's website while the graph on the right describes the Zook's website. Each edge represents one webpage with the pictures and names of two paired suspects.

The activity on both websites is confidential and the authorities cannot monitor the user on his website. The Yook is informed that a suspect is a Zook if and only if he appears on at least two pages of his website and similarly, the Zook is informed that a suspect is a Yook if and only if he appears on at least two pages of his website and that information is accurate since the authorities are not allowed to cheat.

The authorities are aware of the witnesses' choice procedure. Each witness naively looks for a suspect who belongs to the other group (i.e. his name appears on at least two pages of his website). He can use a search engine in the process. A witness starts by checking the group identity of the actual terrorist and in case he does not belong to the opposite group (i.e. his name appears only once in the website), he repeats the process starting with the other name that appears with the terrorist on the only page that the terrorist's name and picture are displayed.

In order to demonstrate the authorities' scheme, assume that the terrorist is z_1. The Yook starts by searching for the name z_1 and finds that it appears on four pages, namely that the terrorist is a Zook and thus he can happily announce that he has identified z_1 to be the terrorist. The Zook also starts by searching for the name z_1 and discovers that on his website the name appears only on one page, together with y_2. He concludes that he has to cheat. Hoping that y_2 appears on at least one more page, he continues by searching for y_2 and finds that it appears on his website more than twice and thus y_2 is a Yook. He then announces that the terrorist is y_2. Thus, both witnesses identify a suspect from the other group.

The DrSeussLand authorities now have two identifications. Nevertheless, they conclude that z_1 is the terrorist. Their logic is based on an understanding of the choice procedure in such an environment. They know that if the terrorist is z_1 (as the Yook claims), the choice procedure would indeed lead the Yook to claim that the terrorist is z_1 and the Zook to claim that the terrorist is y_2. They also know that if the terrorist is y_2 (as the Zook claims), then the Yook witness would announce z_2 rather than z_1. This is because he would start from y_2, and after discovering that he is not a Zook, would continue by searching for z_2 (suspect y_2's partner on the only page in the Yook's website where y_2 appears). The asymmetry leads the authorities to conclude that the terrorist is z_1 and not y_2. It is easy to verify that the above magic would work in the case that any other suspect is the terrorist.

3.2 *The general case*

The implementation problem: The set of states S is finite and is partitioned into W^1 and W^2, each of which has at least two states. Recall our assumption that both parties receive complete information about the state.

The language: A text is characterized by a set D of doubletons of states in S and has the following frame:

$T(D)$: You deserve winning the prize if the state of nature is a member of at least two doubletons in D.

The interpretation of $T(D)$ is the set of all states that belong to at least two sets in D.

The choice procedure: A party is endowed with a technology that provides him with answers to questions of the type $Q(D,s)$: "Which sets in D contain s?" Denote by $A(D,s)$ the answer to question $Q(D,s)$ which is either:

(i) none;

(ii) one doubleton $\{s,t\}$ (which contains s); or

(iii) a set of at least two doubletons (each containing s).

We assume that a party that receives a text $T(D)$ and observes the state s activates the following choice procedure:

Start by asking the question $Q(D,s)$. If $\lvert A(D,s)\rvert > 1$, then declare s. If $A(D,s)$ contains only $\{s,t\}$, then ask the question $Q(D,t)$. If $\lvert A(t)\rvert > 1$, then declare t. Otherwise, that is if $A(D,s) = \emptyset$, or, $A(D,s) = A(D,t) = \{\{s,t\}\}$, repeat the process starting with an arbitrary state x for which the question $Q(D,x)$ was not asked previously. If you have exhausted all states in S without finding an element that belongs to two sets with labels in D, then give up and declare s.

Notice that unless no state satisfies the text $T(D)$ this procedure will always end up with the party finding a state that satisfies it. This is the reason why given this section's framework, the principal cannot achieve his goal with a single agent (unless he always or never deserves the prize).

The choice procedure induces the relation $\rightarrow_{T(D)}$ defined by $s \rightarrow_{T(D)} t$ if either:
(i) $A(D,s)=\{\{s,t\}\}$ and $|A(D,t)|>1$; or
(ii) $|A(D,t)|>1$ and either "$A(D,s)=\emptyset$" or "$A(D,s)=\{\{s,x\}\}$ and $|A(D,x)|=1$".
In option (i), the party asks the question $Q(D,s)$ and discovers that s appears only in $\{s,t\}$. He is then nudged to ask $Q(D,t)$ and discovers that t satisfies the text.
In option (ii), the party starts with $Q(D,s)$ and then is stuck, either because s does not appear in any doubleton in D, or it appears only once with another state that also appears only once. In either of these cases, the party picks an arbitrary state not explored before. Eventually, he must find a state that satisfies the text.

Example: Consider the problem $\langle S=\{1,2,3,4\}, W^1=\{1,2\}, W^2=\{3,4\}\rangle$.
Let $D^1=\{\{1,2\},\{1,3\},\{2,4\}\}$. The text $T(D^1)$ is solved by 1 and 2 only. If the state is 3, then party 1 will declare the state 1 and if the state is 4, then he will declare the state 2. That is, $3 \rightarrow_{T(D^1)} 1$ and $4 \rightarrow_{T(D^1)} 2$.
Let $D^2=\{\{1,4\},\{2,3\},\{3,4\}\}$. Similarly, the text $T(D^2)$ is solved by 3 and 4 only and induces the relation $1 \rightarrow_{T(D^2)} 4$ and $2 \rightarrow_{T(D^2)} 3$.
Clearly, the pair of texts $T(D^1)$ and $T(D^2)$ magically implements the problem.

The following claim generalizes the example:

Claim A *Let $\langle S, W^1, W^2\rangle$ be a problem satisfying that every W^i contains at least two states. If the principal is equipped with the above language and both parties use the above procedure, then the problem is magically implementable.*

Proof. Enumerate the states of W^1 and W^2 as $z^1,\dots,z^K$ and $y^1,\dots,y^L$, respectively (without loss of generality assume that $K \le L$). Form a sequence $x^1,\dots,x^{2L}$ starting with y^1 and alternating between states in W^1 and states in W^2 (for the the case $K=L$, we denote $x^{2L+1}=x^1$). The states in W^2 appear in order, i.e. $y^1,\dots,y^L$. The states of W^1 appear cyclically (if necessary) in their order, i.e. $z^1,\dots,z^K$. Thus, for example, if $K=3$ and $L=5$ the sequence will be: $(x^1,\dots,x^{10})=(y^1,z^1,y^2,z^2,y^3,z^3,y^4,z^1,y^5,z^2)$. The key in this construction is that there is no pair of states $z\in W^1$ and $y\in W^2$ such that z appears right after y somewhere in the sequence and appears right before y elsewhere in the sequence.

We now construct two texts $T(D^1)$ (assigned to party 1) and $T(D^2)$ (assigned to party 2). The set D^1 consists of:
(i) K doubletons $\{z^1, z^2\}, \{z^2, z^3\}, \ldots, \{z^K, z^1\}$; and
(ii) $|W^2|$ doubletons $\{x^k, x^{k+1}\}$, one for each $x^k \in W^2$.
For any $z \in W^1$, we have $|A(D^1, z)| \geq 2$ and therefore, there is no y for which $z \rightarrow_{M(D^1)} y$. For any $y \in W^2$, the set $A(D^1, y)$ contains only one doubleton $\{y, z\}$ where z is the state that appears right after y in the sequence $(x^1, \ldots, x^{2L})$, and therefore y does not satisfy $T(D^1)$. This z is in W^1 and $|A(D^1, z)| > 1$. Thus, there is a unique $z \in W^1$ such that $y \rightarrow_{T(D^1)} z$.

Figure 2 above illustrates the construction of the texts $T(D^1)$ (on the left) and $T(D^2)$ (on the right) in the case that $W^1 = \{z_1, z_2, z_3\}$ and $W^2 = \{y_1, y_2, y_3, y_4, y_5\}$.

For any $y \in W^2$, the set $A(D^2, y)$ contains at least two doubletons and therefore there is no z for which $y \rightarrow_{T(D^2)} z$. For any $z \in W^1$, the set $A(D^2, z)$ contains only one set $\{y, z\}$ where y appears right after z in the sequence $(x^1, \ldots, x^{2L})$ and therefore $|A(D^2, z)| = 1$. This y is in W^2 and $|A(D^2, y)| > 1$. Thus, there is a unique $y \in W^2$ such that $z \rightarrow_{T(D^2)} y$.

By the construction of the sequence $(x^1, \ldots, x^{2L})$ there is no case where y comes right after z and also z comes right after y. Thus, for no $z \in W^1$ and $y \in W^2$ do we have both $y \rightarrow_{T(D^1)} z$ and $z \rightarrow_{T(D^2)} y$. It follows then that $(T(D^1), T(D^2))$ magically implements the problem. ■

The intuition underlying the construction of the two texts is as follows: If $s \in W^i$, then party i will find that at least two doubletons in D^i contain s and he will declare s. Party j starts with s and finds that only one doubleton $\{s, t\}$ is in D^j. He then asks about the state t and ascertains that it belongs to two doubletons in D^j and declares t. But party j has fallen into a trap! The principal can apply the honest-cheater asymmetry principle (given the state s, party j may cheat using t and if the state were t, then party i could not cheat by declaring s) and concludes that j is the cheater.

4. Setting a logical riddle

This section investigates the concept of magical implementation when the language of the principal and the parties' choice procedure are similar to those discussed in Glazer and Rubinstein (2012) in the context of a *single*-agent implementation problem.

4.1 A motivating example

The rector of a university, a magician in his spare time and known for his eccentric academic opinions, is consulting with two professors, *Plus* and *Minus*, about the appointment of Professor G to a prestigious university chair. *Plus* and *Minus* have each interviewed G. Although they agree on G's academic merits, *Plus* firmly supports the appointment, while *Minus* is vehemently against it.

The rector will form his opinion on the basis of the truth or falseness of three statements about G: "he is a genuine scholar in Economics", "he is a genuine scholar in Law", and "he is a genuine scholar in Psychology". The professors do not know which combinations of the facts will bring the eccentric rector to favor the appointment.

The rector could simply ask the professors to state the facts (about which they agree, as mentioned) but he fears that they may give him a biased opinion due to their strong personal bias for or against G. He suspects that they may not be above deception in order to "save the university from what they consider to be a catastrophic decision".

The rector meets *Plus* and *Minus* separately and asks them whether each of the following statements is True or False: $E=$"G is a genuine scholar of Economics", $L=$"G is a genuine scholar of Law", and $P=$"G is a genuine scholar of Psychology". Coding "True" as 1 and "False" as 0, there are 8 possible configurations of answers (states): 000, 001, 011, 100, 101, 110 and 111 (for example, 101 stands for E and P being true and L being false).

The rector is not a fan of multidisciplinary experts and he strongly believes that the chair holder should have only one specialization. Thus, he considers the combinations 100, 010, and 001 as necessary and sufficient for G to be appointed.

Before asking the three questions, the rector (who, as you will recall, is a magician in his spare time) provides each of the professors with a text consisting of a list of propositions that truthfully characterizes the states in which the rector's view is aligned with the professor's. He can do this in many ways, and as we will see, his design of the texts will take into account his knowledge of how each professor will respond to a text. Namely, will he tell the truth, or will he lie? And if he does lie, in what manner will he do so?

The propositions in each of the texts take the form $A \wedge B \rightarrow C$, interpreted as "if the facts in the antecedent, i.e. A and B, are true in the case of G, then the fact in the consequent, i.e. C, must also be". For instance, the first proposition on *Plus*' list is "if G is not a scholar in Economics and is a scholar in Law, then he must not be a scholar in Psychology."

You are right if your report satisfies the following propositions:	
Professor *Plus*	Professor *Minus*
$\neg E \wedge L \rightarrow \neg P$	$L \wedge \neg P \rightarrow E$
$E \wedge L \rightarrow P$	$\neg E \wedge P \rightarrow L$
$E \wedge P \rightarrow \neg L$	$\neg L \wedge \neg P \rightarrow \neg E$
$E \wedge \neg L \rightarrow \neg P$	
$\neg E \wedge \neg L \rightarrow P$	

Figure 3. Each node in the lefthand cube represents a truth configuration (a state). The red dots denote the states that support *Plus*' position, while the black dots denote those that support *Minus*'. The propositions on *Plus*' (*Minus*') list are represented by red (black) arrows in the righthand cube. For example, the red arrow from 101 to 100 represents the proposition $E \wedge \neg L \rightarrow \neg P$ on *Plus*' list.

As mentioned, the rector knows the choice procedure of the professors after they receive such texts. The procedure is recursive. It starts by determining whether the truth satisfies all the propositions on the list. If it does, then the professor announces the truth; if it does not, then he modifies the true state by switching the truth value of the consequent in the proposition that the true state violates and then proceeds from there as if it is the truth. He continues until he reaches a state that satisfies all the propositions on his list.

Assume, for example, that G is an expert in Economics (E) and Law (L), but not in Psychology (P). The state $E, L, \neg P$ (i.e., 110) satisfies all three propositions on *Minus*' list and therefore he reports the truth to the rector. This state violates the proposition $E \wedge L \rightarrow P$ on *Plus*' list, leading him to consider the state E, L, P (i.e., 111), which violates a different proposition on his list, namely $E \wedge P \rightarrow \neg L$. As a result, he then considers the state $E, \neg L, P$ (i.e., 101), but it also violates a different proposition on his list, namely $E \wedge \neg L \rightarrow \neg P$. He finally considers the state $E, \neg L, \neg P$ (i.e., 100), which satisfies all the propositions on his list.

The rector faces two opposing statements, each corresponds to the view of the Professor who made the statement. He decides against the nomination based on the following reasoning: The statements of both professors are the outcomes of their respective choice procedures if *Minus*' statement, $E, L, \neg P$, is true. On the other hand, if the true state is $E, \neg L, \neg P$, then *Plus* would indeed declare that state but *Minus* would declare $\neg E, \neg L, \neg P$ (since the state $E, \neg L, \neg P$ violates the proposition $\neg L \wedge \neg P \rightarrow \neg E$ on *Minus*' list and as a result he checks the state $\neg E, \neg L, \neg P$ which indeed satisfies all the propositions on his list and therefore he declares that state) instead of what he actually declared. Thus, the rector concludes that it is more likely that *Minus*' declaration is correct.

It is easy to verify that the same logic would lead the rector to make a correct inference in each of the other states as well.

4.2 The general case

The implementation problem: Let $S = \{0,1\}^K$ where $K \geq 3$. A state $x_1 x_2 \ldots x_K$ (shortened for $(x_1, \ldots, x_K)$) is a vector representing the truth values of the binary variables $v_1, \ldots, v_K$, where $x_k = 1$ indicates the "truth" of the variable v_k and $x_k = 0$ indicates its "falsity". Recall that we assume that both parties receive complete information about the state. Two states s and t are neighbors, denoted by sNt, if they differ in exactly one component. Assume that each of the sets W_1 and W_2 contains at least two states.

The language: A *text* is characterized by a set of propositions in propositional logic, Φ, each of which uses some of the variables $v_1, \ldots, v_K$ and has the structure $\wedge_{v \in V} \phi_v \rightarrow \phi_z$

where V is a non-empty subset of variables, z is a variable that is not in V, and every ϕ_v is either v (the variable v) or $\neg v$ (the negation of v) and ϕ_z is either z or $\neg z$. For any such proposition ϕ, denote by $\alpha(\phi)$ the antecedent of ϕ and by $\beta(\phi)$ the consequent of ϕ (that is, $\phi = \alpha(\phi) \rightarrow \beta(\phi)$). For example, $\phi = v_1 \wedge \neg v_2 \rightarrow \neg v_4$ is such a proposition; its antecedent is $\alpha(\phi) = v_1 \wedge \neg v_2$ and its consequence is $\beta(\phi) = \neg v_4$. Such a proposition should be interpreted as follows: "If the state satisfies the antecedent of the proposition, then it should also satisfy its consequent." A proposition is *complete* if all K variables appear in it. A complete proposition excludes one state that satisfies the antecedent but not the consequent. For example, when $K = 3$, the complete proposition $v_1 \wedge \neg v_2 \rightarrow \neg v_3$ excludes the state $(v_1, v_2, v_3) = (1, 0, 1)$.

Finally, a *text* $T(\Phi)$ has the following form:

$T(\Phi)$: You deserve the prize if the state satisfies all the propositions in Φ.

The interpretation of a text $T(\Phi)$ is the set of all states that satisfy all the propositions in Φ. For example, if $K = 3$ and Φ contains the five propositions on *Plus*' list, then $T(\Phi)$ contains 100, 010 and 001.

One additional constraint on a text is the condition of *coherence* described in Glazer and Rubinstein (2012): the text should not include two propositions such that their antecedents do not contradict (i.e., no variable v appears in the antecedents once as v and once as $\neg v$), but their consequents do (i.e., the same variable z appears in the consequents of both propositions – in one case as z and in the other as $\neg z$). For example, a text that includes two propositions, $v_1 \rightarrow v_3$ and $v_2 \rightarrow \neg v_3$, is not coherent.

The choice procedure: A party that receives a text $T(\Phi)$ and is informed that the true state is s^* examines states sequentially until either he finds a state that satisfies the text, or he runs out of states to examine. After reaching some state s, he might examine another state t if:

(i) s does not satisfy at least one of the propositions in Φ; and

(ii) for every variable v for which s and t differ, there is a proposition $\phi \in \Phi$ such that s and t satisfy $\alpha(\phi)$ and s does not satisfy $\beta(\phi)$ while t does.

When s and t satisfy the above, we write $s \rhd_{T(\Phi)} t$.

The choice procedure executes the following algorithm starting with the true state:

Set $s = s^*$.
Step 1. Determine whether s satisfies all propositions in Φ.
If it does, then announce it.
If it does not, then proceed to Step 2.
Step 2. Find a new state, t, which was not examined earlier, such that $s \rhd_{T(\Phi)} t$.
If you do not find such a state, then announce s.
If t exists, then go back to Step 1 with $s = t$.

The procedure induces a binary relation $\rightarrow_{T(\Phi)}$ on S defined by $s \rightarrow_{T(\Phi)} t$ if there is a sequence of states $s_1 = s, s_2, \ldots, s_m = t$ such that $s_1, \ldots, s_{m-1}$ don't satisfy $T(\Phi)$, t does satisfy $T(\Phi)$ and $s_l \rhd_{T(\Phi)} s_{l+1}$ for $l = 1, \ldots, m-1$.

The following claim shows that every implementation problem is magically implementable (recall that we assume that $K \geq 3$ and both W^1 and W^2 include at least two states).

Claim B *If the principal is equipped with the above language and both parties use the above choice procedure, then any problem $\langle S = \{0,1\}^K, W^1, W^2 \rangle$ is magically implementable.*

Proof. Recall that a Hamiltonian cycle of the set S is an enumeration $x_1, \ldots, x_{2^K}$ of all states of S such that $x_k N x_{k+1}$ for all k (and $x_{2^K} N x_1$).

Lemma: There is an Hamiltonian cycle with more than one block of W^1-states (and thus also more than one block of W^2-states).
Proof: For every $K \geq k \geq 1$ and $\delta \in \{0,1\}$, let $S_{k,\delta}$ be the set of all states s such that $s_k = \delta$. Let k be a dimension for which both $S_{k,0}$ and $S_{k,1}$ include an element of W^1. Either $S_{k,0}$ or $S_{k,1}$ contains two states of W^2, unless W^2 has exactly two states, one in $S_{k,0}$ and one in $S_{k,1}$, in which case W^1 has at least two elements in one of the sets. Therefore, we can assume without loss of generality that $S_{1,1}$ contains at least two states from W^1 and one from W^2 and there is a state $t \in S_{1,0} \cap W^2$.

Construct an arbitrary Hamiltonian cycle of $S_{1,1}$. If it contains more than one W^1-block we can extend it to an Hamiltonian cycle of S with more than one W^1-block. If not, there must be two W^1-successive states a and b, such that b comes right after a in the cycle. Construct a Hamiltonian cycle of S as follows: start with b, continue in the $S_{1,1}$-cycle up to a, move to a's neighbor in $S_{k,0}$ and continue with an Hamilotian cycle of $S_{k,0}$ that ends with b's neighbor in $S_{k,0}$. This Hamiltonian cycle contains at least two blocks of W^2-states. Figure 4 demonstrates the construction:

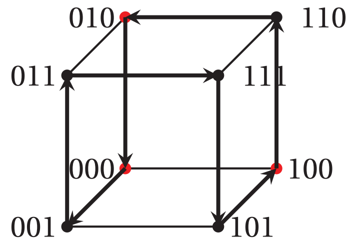

Figure 4. The construction of the Hamiltonian cycle with at least 2 blocks of each color ($a = 110$ and $b = 111$).

Let $x_1, \ldots, x_{2^K}$ be an Hamiltonian cycle of S that has at least two blocks of W^1-states (and thus also of W^2). Let $\to$ be the binary relation on S defined by $s \to t$ if t appears just after s in the cycle. Define $s \to^1 t$ if $s \to t$ and $s \in W^2$ and $s \to^2 t$ if $s \to t$ and $s \in W^1$.

Given any two neighboring states s and t, let $\phi_{s,t}$ be the complete proposition that is satisfied by t but not by s. In other words, given that $s_k \neq t_k$:
(i) the consequent of $\phi_{s,t}$ is v_k if $t_k = 1$ and $\neg v_k$ if $t_k = 0$; and
(ii) the antecedent of $\phi_{s,t}$ is a conjunction of $K-1$ variables or negations of variables, such that for every $l \neq k$ there is one element in the conjunction, either v_l if $s_l = t_l = 1$ or $\neg v_l$ if $s_l = t_l = 0$.

Finally, let Φ^i be the set of all propositions $\phi_{s,t}$ for which $s \to^i t$. The set of states that satisfy the text $T(\Phi^i)$ is exactly W^i. Obviously, the texts are coherent. It is left to show that the pair of texts $(T(\Phi^1), T(\Phi^2))$ magically implements $\langle S, W^1, W^2 \rangle$.

Consider, for example, $s \in W^1$. The state satisfies $T(\Phi^1)$ and thus party 1 declares s. Party 2 finds that s does not satisfy $T(\Phi^2)$. The choice procedure leads him to declare the first $t \in W^2$ in the first W^2-block that follows (in the cycle) the W^1-block that contains s.

The principal can now apply the honest-cheater asymmetry principle and will conclude that 2 is cheating. If the true state were t (as 2 claims), then party 1 would declare state r, the first element in the W^1-block that follows the W^2-block that contains t. Since there are at least two W^1-blocks, the state r is different from the state s. ■

We now contrast the model with the single-agent model studied in Glazer and Rubinstein (2012). The set of the states and the principal's language are the same. The requirement here that the implementation should be by means of two texts that fully describe the circumstances under which each party deserves the prize is in the spirit of our notion there of "truthful implementation". One difference is in the choice procedure. There, we assumed that if the truth s^* does not satisfy the text $T(\Phi)$, then the agent examines a modified state t such that $s^* \rhd_{T(\Phi)} t$ and if t does not satisfy $T(\Phi)$, then the decision maker returns to s^* and looks for other modifications triggered by propositions in Φ violated by s^*. In contrast, here we assume that the decision maker continues recursively with t.

The current choice procedure was mentioned in Glazer and Rubinstein (2012) and we know from there that a set of states W is truthfully implementable given this choice procedure if and only if each connected component (with respect to the neighboring binary relation) of the complementary set of W contains a cycle of length 4 or more. Thus, if either W^1 or W^2 has this feature, then implementation can be accomplished by involving only one agent. However, as in the motivating example in subsection 4.1., this is often not the case and implementation requires the involvement of both parties.

Notice that the conclusion of claim B remains valid if we assume that the parties follow the choice procedure analysed in our previous work.

5. Making cheating too risky

The motivating example in this section is the two invigilators scenario presented in the introduction. If the student received the actual answer to the question, then L (who likes the student) would like to claim that the student only received a recommendation to start with a particular equation. But such a claim would require L to guess the equation that the whispered message "$\alpha = 3$" refers to and unlike the principal, L is not familiar with the exam and by making such a claim he would take the risk of being caught cheating.

Conversely, if the student merely received a recommendation to start with a specific equation, then H (the hostile invigilator) is not taking any risk by cheating, i.e. he can solve the equation and claim that the student received the actual solution to the equation. This asymmetry enables the principal to infer that the student did not receive the actual answer to the question when one invigilator claims that the student received the answer while the other claims that the student only received a recommendation to start with a particular question that indeed appears in the exam.

We now formalize the example within a broader framework.

The implementation problem: We expand the notion of an implementation problem to a tuple $\langle S, W^1, W^2, I, \mu, I_p \rangle$ where the three additional elements are:

I: An information structure of S (a partition of S). In state $s \in S$, each party is informed about the cell $I(s)$ in the partition that contains s.

μ: A probability measure on S.

I_p: The principal's information partition of S. In state $s \in S$, the principal is informed about the cell $I_P(s)$ in the partition that contains s.

It is required that any information set in I is either a subset of W^1 or of W^2, that is, the information possessed by the parties is sufficient to determine which party deserves the prize. This assumption is not imposed on I_p since the principal would then be able to make the correct decision without eliciting any information from the parties.

The language: A text has two parameters: Y which is a subset of S and constitutes the information provided to the parties by the principal (in addition to what the parties already know according to the information structure I) and a set $W \subseteq S$, which is a union of cells in I and is interpreted as a description of the cells in which the party that receives the text deserves the prize. A text $T(Y,W)$ has the following form:

$T(Y,W)$: The state of nature is in Y. You deserve the prize if the state is in W.

The choice procedure: The choice procedure in this section is more conventional than in the previous two. After receiving a certain information set, a party's willingness to cheat by reporting a false information set depends on his fear of getting caught. We say that a party is *caught cheating* if his announcement and the principal's knowledge do not intersect. It is assumed that a party believes that the information he receives from the principal is truthful (but does not necessarily consist of all the information possessed by the principal). It is further assumed that a party considers cheating (by announcing an untrue information set in I) only if he believes (given the prior, the information set in I which the parties initially received, the principal's announcement and the principal's information structure I_p) that the probability of getting caught does not exceed some threshold τ.

To summarize, the procedure followed by a party after receiving the text $T(Y,W)$ and given that he initially received the information set $K \in I$ is as follows:

If $K \subseteq W$, then declare K. If $K \nsubseteq W$, then search for an $L \in I$ in which you deserve the prize ($L \subseteq W$) and the probability of being caught after declaring L is below τ, that is, $\mu(\{s \mid L \cap I_p(s) = \emptyset\} \mid K \cap Y) \leq \tau$. If you find such an L, then declare it; otherwise declare K.

The procedure generates the binary relation $\rightarrow_{T(Y,W)}$ on the information sets in I, defined by $K \rightarrow_{T(Y,W)} L$ if $K \nsubseteq W$, $L \subseteq W$, and a party that initially receives the information K and cheats by declaring L gets caught with probability (conditional on $K \cap Y$) not exceeding τ.

Finally, we need to modify the definition of magical implementation. The principal chooses a a partition I_p^*, one that is coarser than I_p, and after being informed about an information set M in I_p he provides each party i with the text $T(Y, W^i)$ where Y is the cell in I_p^* that includes M. Thus, the principal always provides each party i with the truthful description of the circumstances in which he deserves the prize (the set W^i) and the information about the state (the set Y) is always the truth, but not necessarily all of it. In order to magically implement his goal, it is required that if the parties follow the choice procedure, then whenever their claims differ the principal will be able to activate the honest-cheater asymmetry principle and correctly identify the deserving party.

Formally, we say that $\langle S, W^1, W^2, I, \mu, I_p \rangle$ is magically implementable if there is a partition I_p^* coarser than I_p such that for any set $Y \in I_p^*$ the texts $T(Y, W^1)$ and $T(Y, W^2)$ satisfy that if $K, L \in I$, $K \cap Y \neq \emptyset$, $L \cap Y \neq \emptyset$ and $L \rightarrow_{T(Y,W^i)} K$, then $K \nrightarrow_{T(Y,W^j)} L$.

Notice the difference between catching a cheating party and inferring that a party is cheating using the honest-cheater asymmetry principle. The former occurs only when the principal has solid proof that a party is cheating, namely the party's statement contradicts the information he possesses. The latter occurs when there is a dispute between the parties and the principal applies the honest-cheater asymmetry principle *without any solid proof* that one of the parties is cheating.

The exam example: We now formalize now the example from the introduction. The set S consists of four states depicted as cells in the table below, where a row indicates the content of the whispered message (the solution or just the equation) while a column represents the equation (assuming, for simplicity, that there are only two possible equations):

the whispered message	The exam's equation	
	$\alpha+1=4$	$\alpha+2=5$
a solution	a	b
an equation	c	d

Each party's information partition is $I = \{\{a, b\}, \{c\}, \{d\}\}$ and the principal's information partition is $I_p = \{\{a, c\}, \{b, d\}\}$. We assume that $\mu(a) = \mu(b) > 0$. The winning sets are $W^H = \{a, b\}$ and $W^L = \{c, d\}$.

When $\tau < 1/2$, the texts $T(S, W^H)$ and $T(S, W^L)$ (the principal does not provide any additional information to the parties) magically implement the problem:
– When the parties are informed that the state is c (d), they know that the principal possesses the information $\{a, c\}$ ($\{b, d\}$). Then, H is not afraid to cheat and declare $\{a, b\}$ since with certainty he will not be caught cheating.
– When the parties are given the information $\{a, b\}$, they do not know whether the principal received the information $\{a, c\}$ or the information $\{b, d\}$. Party L is afraid to report $\{c\}$ since if he does, he will be caught cheating with probability $\mu(b)/[\mu(a)+\mu(b)] = 1/2 > \tau$. Similarly, L is afraid to cheat by reporting $\{d\}$.

Thus, $\{c\} \to_{T(X,W^H)} \{a, b\}$ and $\{d\} \to_{T(X,W^H)} \{a, b\}$ but $\{a, b\} \not\to_{T(X,W^L)} \{c\}$ and $\{a, b\} \not\to_{T(X,W^L)} \{d\}$ and the problem is magically implementable.

Note that if $\tau > 1/2$, then we will also have $\{a, b\} \to_{T(X,W^L)} \{c\}$ and $\{a, b\} \to_{T(X,W^L)} \{d\}$ and the pair of texts fails to magically implement the problem. This will also be the case if the principal reveals his knowledge. Then, in state a (for example), the principal will announce $\{a, c\}$ and L (who possesses the information $\{a, b\}$) will not be afraid to cheat by declaring $\{c\}$. At the same time, H will not be afraid to cheat in state c by declaring $\{a, b\}$. Thus, full information revelation will not enable the principal to perform his magic.

The asymmetry between a cheater and a truth-teller arose on its own in the above example. We will see that in the general framework, the principal may wish to manipulate the situation by sharing some of his information with the parties in order to create asymmetry between a truth-teller and a cheater.

Setting a trap by providing additional information: The following example demonstrates that providing information is sometimes necessary for magical implementation:

$S = \{a, b, c, d, e, f\}$,

$W^1 = \{a, b, c\}$, $W^2 = \{d, e, f\}$,

$I = \{W^1, W^2\}$, $\mu =$ the uniform probability measure on S,

$I_p = \{I_1 = \{a\}, I_2 = \{c, d\}, I_3 = \{b, e\}, I_4 = \{f\}\}$ and

$1/3 = \mu(a)/\mu(W^1) < \tau < \mu(a)/[\mu(a)+\mu(c)] = 1/2$.

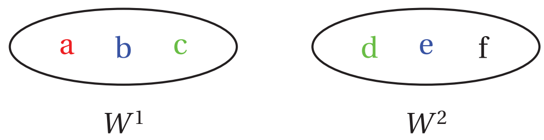

Figure 5. The information partitions I is represented by the ellipses while I_p is represented by the colors.

Magical implementation is not possible without providing additional information, since both $W^1 \to_{T(S,W^2)} W^2$ and $W^2 \to_{T(S,W^1)} W^1$. That is, in W^1 party 2 is not deterred from announcing W^2 and in W^2 party 1 is not deterred from announcing W^1.

Magical implementation is obtained by the principal committing to supply additional information according to the information structure $I_p^* = \{I_1 \cup I_3, I_2 \cup I_4\}$. Assume that the principal has announced $I_1 \cup I_3$ (or analogously $I_2 \cup I_4$). In the case that they are initially informed of W^1, the parties will conclude that the state of nature is either a or b. Party 2 is deterred from cheating since he will be caught with probability $1/2 = \mu(a)/[\mu(a)+\mu(b)] > \tau$. In the case that the parties are informed of W^2, they will conclude that the state of nature is e and therefore the principal possesses the information $\{b, e\}$ and party 1 will not be caught cheating if he announces W^1. Thus, $W^1 \nrightarrow_{T(I_1 \cup I_3, W^2)} W^2$ and $W^2 \to_{T(I_1 \cup I_3, W^1)} W^1$ and in the case of disagreement the principal will be able to use the honest-cheater asymmetry principle to infer that party 1 is the cheater.

6. Comments by Ariel Rubinstein

a. I am fully aware (and proud) that the paper is written in a style different from what is the convention these days in Economic Theory. The discussion is purely conceptual. We do not claim that there are any applications. The paper is short. Although the discussion is carried out in formal language we avoid any fancy mathematics. The goal is simply to convey an idea. Indeed, I am suspicious of any work in Economic Theory that goes beyond presenting one main idea with a few simple examples. This paper should be read almost like a story: you might find it interesting, entertaining or elegant, or maybe ... not. If the reader derives something useful from the article, that's fine; however, it is not my intention to generate any "practical" conclusions.

b. The paper contributes to "implementation theory", although we diverge from the traditional commitment to game-theoretical tools. While we of course acknowledge the appeal of game-theoretical models – having employed them ourselves in the past – we wish to challenge the status of game theory as the exclusive approach used in the implementation literature.

c. I doubt that people in the real world use any of the described procedures of cheating "as is". In Glazer and Rubinstein (2012), we claim – quite convincingly if I may say so myself – that hints of the procedure described there have been observed in experimental data. Nevertheless, we have refrained from running another set of experiments here. Such experiments might have been interesting, but they are not necessary in order to construct interesting stories.

References

Ben-Porath Elchanan, Eddie Dekel and Barton L. Lipman (2019). "Mechanisms with Evidence: Commitment and Robustness". *Econometrica, 87*, 529-566.

Glazer, Jacob and Ching-To Albert Ma (1989). "Efficient Allocation of a 'Prize' — King Solomon's Dilemma". *Games and Economic Behavior, 1,* 222-233.

Glazer, Jacob and Ariel Rubinstein (2012). "A Model of Persuasion with Boundedly Rational Agents". *Journal of Political Economy, 120*, 1057-1082.

Glazer, Jacob and Ariel Rubinstein (2006). "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach". Theoretical Economics, 1 (2006), 395-410.

Green Jerry R. and Jean-Jacques Laffont (1986). "Partially Verifiable Information and Mechanism Design". *Review of Economic Studies, 53*, 447-456.

Lipman, Barton L. and Duane J. Seppi (1995). "Robust inference in communication games with partial provability". *Journal of Economic Theory, 66*, 370-405.

Maskin, Eric (1999), "Nash Equilibrium and Welfare Optimality". *The Review of Economic Studies, 66*, 23-38.

Michelbacher, Lukas, Stefan Evert, and Hinrich Schütze (2007). "Asymmetric Association Measures." *Proceedings of the 6th International Conference on Recent Advances in Natural Language Processing,* 367-372.

Motty Perry and Philip J. Reny, (1999). "A General Solution to King Solomon's Dilemma." *Games and Economics Behavior, 26*, 279-285