PRINCIPAL-AGENT PROBLEMS FROM A GAME-THEORETIC VIEWPOINT

Martin SUMMER

Forschungsbericht/
Research Memorandum No. 347

July 1994

Abstract:

This paper discusses the game-theoretic foundations of the famous Principal-Agent model. Taking the game-theoretic viewpoint allows to bring the different versions of this model into a unified perspective. Since the Principal-Agent model deals with situations, where some players have information that others have not, some form of communication is necessary for the exchange of information. A central message of the model is that for the exchange of information players have to be provided with the right incentives. It is shown how the concept of correlated equilibrium can be used to add formally the idea of communication to a game in an analytically tractable way. The power of this approach lies in the fact that it allows to simplify the structure of a possibly very complicated game with communication to a problem that is much easier to handle. This result known to the literature as the Revelation Principle is discussed. An extensive example at the end of the paper shows possible limitations of relying on arguments in the spirit of the Revelation Principle.

Zusammenfassung:

Die vorliegende Arbeit diskutiert die spieltheoretischen Grundlagen des bekannten Principal-Agent Modells. Diese Perspektive erlaubt es, die verschiedenen Versionen dieses Modells in einen einheitlichen Rahmen zu stellen. Das Principal-Agent Modell befasst sich mit Situationen , in denen manche Spieler Informationen besitzen, die anderen fehlen, was Kommunikation zum Austausch von Informationen erforderlich macht. Eine zentrale Aussage des Modells ist, dass dieser Informationsaustausch entsprechende Anreize erfordert. Das Konzept eines korrelierten Gleichgewichts ermoeglicht es, Kommunikation in einem Spiel in einer handhabbaren Weise zu analysieren. Dieser Ansatz erlaubt es, ein potentiell sehr kompliziertes Spiel auf ein wesentlich einfacheres Problem zu reduzieren. Dieses Resultat, das in der Literatur als Revelation- Principle bekannt ist, wird diskutiert. Ein ausfuehrliches Beispiel zeigt moegliche Probleme von Argumenten im Sinne des Revelation- Principles.

Principal-Agent Problems from a Game-theoretic Viewpoint

Martin Summer
Institute for Advanced Studies*

1 Introduction

This paper was motivated by a course in the economics of information and uncertainty held at IAS in september last term. Since problems of moral hazard and adverse selection belong to the standard body of microeconomic theory, they are intensively discussed in such a course. They are also covered in most major textbooks on microeconomics.[1]

Usually the models are not set up in game theoretic terms although the questions of interest in such a context fit exactly the language of game theory.

In this paper I, therefore, discuss - using standard results of the game theoretic literature - how one can bring principal- agent problems into a game theoretic perspective. In my opinion this viewpoint has some advantages compared to the approach taken by just writing down the appropriate individual rationality and incentive compatibility constraints to formulate these problems as simple constrained optimization problems. First of all, the language of (non-cooperative) game theory forces the modeller to be specific about the rules of the game by it's representation theory. So it is more specific about the institutional setup one has in mind, which gets easily out of sight if one just writes down a static optimization problem. Furthermore game theoretic language forces the modeller to be specific about the particular solution concepts employed to predict the outcome of strategic interaction. In this way it forces to think more about the possible limitations. Finally the game theoretic approach makes clear that the simple mathematical structure we can use to characterize situations of moral

*I have to thank Simon Gaechter, David Kelsey, Arno Riedl and especially Klaus Ritzberger for many helpful hints and comments

[1]See for example Varian [1992], Kreps [1991] to name the most prominent examples.

hazard and adverse selection are a theoretical *result* known to the literature as the *Revelation Principle.* So it makes it more transparent from a theoretical viewpoint why and most of all why we might *not* be confident that we get the solutions we are looking for. Furthermore the game theoretic viewpoint makes transparent how we can model the exchange of information and the communication that is necessary for this coordination in an analytically tractable way.

Following Myerson[2] I present a completely general framework which encompasses the often seperately discussed problems of *moral hazard* and *adverse selection* in one setup and makes clear the general spirit of the arguments used in deriving an equilibrium of a game with this particular structure. The second advantage already alluded to above lies in the fact that, by looking at principal-agent problems from a game theoretic viewpoint, on can better understand the problems of the approach taken by the revelation principle. If we, for instance, calculate an incentive compatible mechanism we construct the game such that one player called the *principal* can make a take-it-or-leave-it offer concerning the rules of the game played by the other players called the *agents.* It is then often conjectured that the players are willing to accept any Nash equilibrium of this game. It should therefore not be astonishing that this approach inherits all the problems we have with the solution concept of a Nash equilibrium. In general there will be several Nash equilibria of the game played by the agents and even worse, if we would apply some selection criterion like e.g. Pareto-dominance[3], we would not get the equilibrium the principal wants to implement via the revelation principle. I discuss an example due to Kreps[4] and extended by Repullo[5] to make this point clear. In this example I also show that it can be possible to implement a truth-telling equilibrium via an indirect mechanism although the direct mechanism wouldn't do that. In this sense it is not quite correct to say that without loss of generality we can use the revelation principle to characterize equilibria we are interested in. In fact the revelation principle only gives a necessary condition for implementation. To my understanding problems like this become clear only from a game-theoretic perspective. The main theme of the paper is to give a precise argument why one is allowed to use the approach we know from standard texbooks or learn in courses about information economics and why we need to take care what conclusions we draw from applying the standard tools. In this sense the argument is mainly of theoretical interest. All the results and ideas stated here are not new and insofar this paper is no original contribution to the

[2]Myerson[1982],[1991]

[3]See Harsanyi, Selten [1988]

[4]Kreps [1990], p 681 ff.

[5]Repullo [1986]

literature on mechanism design and agency theory. It serves mainly as a personal summary of an approach that seemed most insightful to me when studying this literature.

The first part of the paper introduces the concept of a so- called *game of incomplete information.* This concept gives us a unified framework to represent all situations of asymmetric information that we might be interested in. If people try to exchange information they need some form of communication. The next part of the paper shows how one can add formally the idea of communication to a game with incomplete information in an analytically tractable way. This part of the paper gives also a general formulation of the revelation principle and a proof along the lines of Myerson[6] which treats moral hazard and adverse selection in one setup and makes thereby clear what is the general idea behind these kinds of arguments. The last section gives an example that shows possible limitations of relying on arguments in the spirit of the revelation principle.

2 Games of Incomplete Information

The principal-agent literature[7] focuses it's attention on situations where some agents have information that others have not. Most frequently the analysis concentrates on situations where agents hold private information about decision domains or private information about particular characteristics. Of course a mixture of both might occur in all kinds of applications of interest.[8] In the following I want to introduce the basic game theoretic notions to describe all possible situations of asymmetric information in a unified framework.

If a game starts in a situation where from the very beginning some players have private information about the game not known to all the others we say that we have a game of *incomplete information.*[9] For example one could think of the famous job market signalling game[10] where a worker holds some private

[6]Myerson [1991]

[7]The literature focusing on these problems is so huge that an enumeration would be longer than the paper itself. So I only give references to papers I explicitly refer to.

[8]See, for example, Antle [1982.]

[9]This notion is to be distinguished from a game with *imperfect information.* Imperfect information means that the game contains information sets that have more than one decision node. By incomplete information one means that the players have different information about the circumstances under which a game is played. As will be argued the modelling strategy consists in transforming an incompete information game into a game that can formally be analyzed like a game with imperfect information

[10]Spence [1973]

information about his ability not known to the employer. In the terminology of game theory the private information a player can have is called his *type*.

The original contribution how one can handle situations of incomplete information in a game theoretic framework is by Harsanyi [11]. He proposed a way of representing and solving games of this kind, which is basically a generalization of the concept of a game in strategic form, called *Harsanyi form* or *Bayesian form.*

The formal objects that characterize a game in Bayesian form are a set of players, denoted as N. Every player i in N has a set of possible actions denoted as C_i. The private information a player can hold from the very beginning is described formally as a set T_i for every player i in N. So if we form the respective cartesian products over C_i and T_i, with respect to all players we generate the sets C and T describing the spaces of all possible profiles of actions and types.

The next object we have to introduce to describe a Bayesian game is a probability function p_i, called a belief, that is a mapping from T_i into $\Delta(T_{-i})$, where $\Delta(T_{-i})$ denotes the set of probability distributions over all combinations of possible types other than i. This function is needed to formally describe the nature of private information we were talking about before. If a player i has type t_i, p_i gives him the conditional probability that the others have types that are components of the vector t_{-i}. So for example, $p(t_{-i} \mid t_i)$ is the conditional probability that the actual profile of types is t_{-i} given player i has type t_i.

Each player i in N has preferences over outcomes that are represented as usual by a Von-Neumann-Morgenstern utility function, which is a mapping from $C \times T$ into the real numbers R. If we denote this function as $u_i(c,t)$ we can interpret this number as some VNM utility scale for player i if the actual profile of types is t and the actual profile of actions is c. We call this number player i's payoff.

These objects altogether completely describe a Bayesian game, which is often denoted as

$$\Gamma^b = (N, (C_i)_{i\in N}, (T_i)_{i\in N}, (p_i)_{i\in N}, (u_i)_{i\in N})$$

If the sets N, C_i, T_i are finite for every i, the game is finite.[12]

To have enough information to analyze the game we have to assume that each player knows Γ^b and his own type. This has to be common knowledge among the

[11]Harsanyi [1967], [1968]

[12]Note however that the finitness assumption is not material to the theory and the results discussed here. The continuous case just demands a little bit more of mathematical machinery without adding further insights.

players. [13] The function $p_i(.)$ is a formal expression of the idea that each player tries to guess what the type of all the others might be. Therefore $p_i(.)$ is called player i's belief. In almost all applications in economics one makes the following assumption about $p_i(.)$:

All players i have the same prior distribution $p_i(.)$ on the typespace T.

The usual argument in defense of this assumption called the *Harsanyi Doctrine* is as follows: Since game theory tries to analyze the behavior of rational and intelligent players [14] the beliefs have a certain structure imposed by the rationality postulate. Rational players will necessarily use data in the same way. If two players have different believes they must therefore have different data. Since ex ante the players can't have different data their prior beliefs must be the same, so the players' prior probability measure on T must be the same. If players are rational in this sense and it holds, as assumed, that $p(.)$ is common knowledge among them they update their beliefes from their common prior by Bayes formula whenever they receive some new information. If believes are formed in this way they are called *consistent.* Note that the common prior assumption does not rule out that players can hold different believes *a posteriori.* What the common prior assumption however does is, that different probabilities reflect differences in information only.[15][16]

Formally we call believes $p(.)$ consistent if and only if there exists a probability distribution in $\Delta(T)$ such that

$$p_i(t_{-i} \mid t_i) = \frac{P(t)}{\sum_{s_{-i} \in T_{-i}} P(s_{-i}, t_i)}, \forall t \in T, \forall i \in N$$

Since the approach introduced by Harsanyi rests essentially on the idea to transform a situation of incomplete information into a situation of imperfect information about types, the solution concept appropriate for this setup has nothing fundamentally new to add to the idea of a Nash equilibrium. Harsanyi defined a Bayesian equilibrium to be any Nash equilibrium of the type-agent representation described above.

Formally this can be expressed in the following way:

[13]The players type is a summary of everything he knows that is not common knowledge

[14]See Myerson 1991: Myerson defines rational as consistent in pursuit of his own objectives, by intelligent he defines that the players understand everything about the structure of the game that a game theorist understands and can therefore make the same inferences

[15]For the common prior assumption see Aumann [1976], [1987], Binmore [1991]

[16]Although the Harsanyi Doctrine is a standard assumption it is controversial. A decision theorist could object that it is an additional assumption to the Savage axioms and if it was purely an implication of rational choice ,as formulated in the argument above, it should be derivable from those axioms. This argument was pointed out to me by David Kelsey.

Definition 1 : (Bayesian Nash equilibrium:) *Consider any Bayesian game as given by:* $\Gamma^b = (N, (C_i)_{i\in N}, (T_i)_{i\in N}, (p_i)_{i\in N}, (u_i)_{i\in N})$ *Let* σ *be a randomized strategy profile for* Γ^b *where* $\sigma = ((\sigma_i(c_i \mid t_i))_{c_i\in C_i})_{t_i\in T_i}$, $i \in N$, $\sigma_i(c_i \mid t_i) \geq 0$, $\forall c_i \in C_i$, $\forall t_i \in T_i$, $\forall i \in N$ *and* $\sum_{c_i\in C_i} \sigma_i(c_i \mid t_i) = 1$, $\forall t_i \in T_i$, $\forall i \in N$. *A Bayesian equilibrium of* Γ^b *is any randomized strategy profile such that for every player* i *in* N *and every type* t_i *in* T_i

$$\sigma_i(. \mid t_i) \in \arg\max_{\tau_i\in\Delta(C_i)} \sum_{t_{-i}\in T_{-i}} p_i(t_{-i} \mid t_i) \sum_{c\in C}(\prod_{j\in N-i} \sigma_j(c_j \mid t_j))\tau(c_i)u_i(c,t).$$

3 Games of incomplete information and communication

Since the principal agent literature has emerged from the analysis of informational problems in economic interactions, the concept of a Bayesian game is the suitable game theoretic setup to study such problems. Two situations have been particularily discussed in the literature. In the first kind of problems the players have private information about some particular characteristic (Adverse selection) in the second class of problems they have private knowledge about their decision domain (Moral hazard).

The exchange of the privately held information among the agents and the principal requires communication. From a theoretical perspective this situation poses the problem how one should include the possibilities of communication into the analytical framework provided by game theory. In principle communication moves could be described by an extensive form. Such a setup however leads to games that could practically be impossible to characterize because every player would have to check every word in the dictionary as a best reply to just one single word said by his opponent.

Another more practicable and useful way is to take communication possibilities into account via the solution concept assuming that to the strategy options explicitly given by the normal form the players have implicit options to communicate with each other. If one takes this idea of implicit communication possibilities seriously on cannot exclude that players would be willing to use correlation devices, which could be thought of as an extended game with some preliminary lottery selecting a vector of signals before the game for every player. Since all communication is thereby conceptually put into the preplay stage the solution concept for such an extended game depends only on it's normal form. This approach is taken by Aumann in his concept of *correlated equilibrium.*[17] So this

[17] Aumann [1974], [1987]

concept could be seen as an approach that treats communication not via the extensive form but via an appropriate solution concept that takes the idea of mediated preplay communication seriously.

It is as well possible to extend the spirit of this approach to Bayesian Games.[18] The idea is to generate a *Bayesian game with communication* from a Bayesian game by adding a general communication device to the game that gets signals as inputs from the players and sends signals as outputs to the players and allows the players to coordinate their strategies by this device. By this step one gets an induced game in normal form for which one can in principle identify the set of Nash equilibria or some refinements of Nash equilibria. The final step is then to show that the set of Nash-equilibria of an induced communication game can be simulated by a representation that is mathematically simpler but essentially equivalent in the sense that the representation can simulate any equilibrium of any game that can be generated from a Bayesian game by adding a communication structure. This kind of argument has become popular in the literature under the name of the *revelation principle*

In the following I want to develope this argument in detail.

3.1 Adding a communication structure to a Bayesian Game

If one thinks of modelling communication by assuming a general communication device that can send recommendations to players as well as receiving inputs from them one can describe such a mediation system or *mechanism* in the following way: Every agent has a set of possible messages that he can receive. Denote this set by M_i. He has also a set of possible reports he can send out to the system. Let the set of possible reports for a player i be denoted by R_i. The respective cartesian products over all players N make up the whole space of messages and reports. A *communication system* can now be defined as a mapping from μ: $R \longrightarrow \Delta(M)$. μ denotes the conditional probability that m would be the vector of messages for all agents if r would be the vector of all reports sent. The characterization of the mapping into all probability distributions over profiles of messages is a formal characterization of the idea that communication can be noisy.

Given such a communication system we can define the set of pure strategies generated from a Bayesian game by adding a communication system μ.

Let

$$B_i = \{(r_i(t_i), \delta_i(m_i, t_i)) \mid r_i : T_i \longrightarrow R_i, \delta_i : M_i \times T_i \longrightarrow C_i\}$$

[18]See especially Myerson [1982],[1991],Forges [1986]

The interpretation of this set is as follows. It gives the pair of reports, that player i can send out into the communication system and actions he can choose as a function of the message or recommendations he gets from the communication system given his type. The action taken by agent i is a function. This can be thought of as giving the players an option to disobey the recommendations they get from the mechanism. The recommendations of the device have no binding force for the players. This expresses the idea that the decision domain for the players is not observable to the principal for instance, so it is not possible to commit to binding contracts. Given any such n-tuple of communication strategies each players expected payoff would be given by:

$$v_i((r_j(.), \delta_j(.))_{j\in N}) = \sum_{t\in T} \sum_{m\in M} p(t)\mu(m(.) \mid r(.))u_i(\delta(m,t),t)$$

This generates the Bayesian game:

$$\Gamma^b_\mu = (N, (B_i)_{i\in N}, (T_i)_{i\in N}, (p_i)_{i\in N}, (v_i)_{i\in N})$$

In principle all that remains to do is to study the Nash equilibria. So we have included communication and can apply a solution concept that depends only on the normal form of the game. Since the message and report spaces are not restricted at all the problem might, however, be very difficult to solve practically.

The revelation principle now tells us that we can reduce any communication system of the kind described above to one with a much simpler structure.

3.2 The revelation principle

To study Bayesian games with communication one starts to look at communication systems with a particular structure. Instead of looking at general message and report spaces we restrict attention to the case where $M_i = C_i$ and $R_i = T_i$. The report space for each player i in N is now identical with his type space and the possible messages or recommendations to act he can get from the system is the set of available actions or pure strategies. We denote then for any c in C and t in T with $\rho(c \mid t)$, $\rho : T \longrightarrow \Delta(C)$ the conditional probability that the strategy profile c would be recommended given that the type profile t was reported. In the terminology of mechanism design such a communication system is called a *direct mechanism.*[19] This mechanism induces in the same manner as described above a Bayesian game with communication. Let

$$B_i = \{(r_i, \delta_i) \mid r_i \in T_i, \delta_i : C_i \longrightarrow C_i\}$$

[19]Dasgupta, Hammond , Maskin [1979]

and the payoff-function belonging to these communication strategies

$$v_i : B \times T \longrightarrow R : v_i((r_j, \delta_j)_{j \in N}, t) = \sum_{c \in C} \rho(c \mid (r_j)_{j \in N}) v_i((\delta_j(c_j))_{j \in N}, t)$$

then:

$$\Gamma_\rho^b = (N, (B_i)_{i \in N}, (T_i)_{i \in N}, (p_i)_{i \in N}, (v_i)_{i \in N})$$

is the game induced by adding ρ to the Bayesian game.

If a player in the game reports his type honestly and obeys the recommendation of the mechanism ρ his VNM-utility would be given by:

$$U_i(\rho \mid t_i) = \sum_{t_{-i} \in T_{-i}} \sum_{c \in C} p_i(t_{-i} \mid t_i) \rho(c \mid t) u_i(c, t)$$

where the minus subscript has the usual meaning of denoting the exclusion of component i. Now since we want to express the idea that there is an asymmetry of information so that recommendations have no binding force and types are private information we must allow the players to report dishonestly and disobey the recommendations of the mechanism. So dishonest reporting and disobedient acting given all the others are honest and obedient is given by:

$$U_i^*(\rho, \delta_i, r_i \mid t_i) = \sum_{t_{-i} \in T_{-i}} \sum_{c \in C} p_i(t_{-i} \mid t_i) \rho(c \mid t_{-i}, r_i) u_i(c_{-i}, \delta_i(c_i), t)$$

Now we can study the Bayesian equilibria of this communication game. One is usually interested in situations where honest reporting and obedient acting is an equilibrium for the players. A communication system or a mechanism is called *incentive compatible* if the following constraint is satisfied:

$$U_i(\rho \mid t_i) \geq U_i^*(\rho, \delta_i, s_i \mid t_i)$$

$$\forall i \in N, \quad \forall t_i \in T_i, \quad \forall s_i \in T_i, \quad \forall \delta_i : C_i \longrightarrow C_i$$

This condition is saying that the expected utility of obedient acting and honest reporting must me greater or equal than the expected utility of disobedient acting and dishonest reporting for all players and all types.

Now we can state the *Revelation principle* (**Myerson 1982**)[20]

Theorem 1 *Given any Bayesian game with communication induced from a Bayesian game by adding a communication system and given any Bayesian equilibrium of the induced game there exists an incentive compatible direct mechanism in which every type of every player gets the same expected utility as in the equilibrium of the induced communication game.*

Proof:

Step 1: Consider any communication system $\mu : R \longrightarrow \Delta(M)$ and take any communication strategy profile $(r_i, \delta_i)_{i=1}^n$. A Bayesian equilibrium for this game is a vector σ that specifies for each i in N, each (r_i, δ_i) in B_i and each t_i in T_i a probability $\sigma_i(r_i, \delta_i \mid t_i)$ that i would report r_i and choose c_i according to $\delta_i(.)$ if his type were t_i.

Step 2: Now we construct an incentive compatible direct mechanism for Γ_μ^b as follows: Let $\delta^{-1}(c, t) = \{m \mid \delta_i(m_i, t_i) = c_i, \forall i\}$. [21] Now construct

$$\rho(c \mid t) = \sum_{(r,\delta) \in B)} \sum_{m \in \delta^{-1}(c,t)} (\prod_{i \in N)} \sigma_i(r_i, \delta_i \mid t_i))\mu(m \mid r) \quad \forall c \in C, \quad \forall t \in T$$

which depends only on c and t as required for an incentive compatible direct mechanism.

Step 3: This mechanism leads to the same probability-distribution over actions given types as the original mechanism. The mechanism constructed this way therefore simulates the outcome that results from equilibrium σ, because $\rho(c \mid t)$ gives exactly the probability of actions c given type t under $\mu(.)$ i.e the probability that players would ultimately use c and report t after participating in the communication system $\mu(.)$.

[20]The content is a standard result known to the literature at least since Gibbard [1973], the formulation and the proof used here is due to Myerson and states the revelation principle in the most general form I found in the literature.

[21]Note that we defined a direct mecanism as a mechanism with $M_i = C_i$ which explains the set $\delta^{-1}(.)$

Step 4: Now it remains to check that the direct mechanism is incentive compatible. Suppose to the contrary that incentive compatibility is violated. Then agent i could gain by reporting s_i if his type were t_i and choosing δ_i^* instead of δ. But then $U^*(\rho, \delta_i, s_i \mid t_i) \geq U_i(\rho \mid t_i)$. By the construction of ρ however this would imply that σ can't be an equilibrium of the original game contrary to our assumption. Hence ρ must be incentive compatible. □

The interesting point of this result lies not so much in it's content because the revelation principle only states that *if* there is an equilibrium of a communication game then you can achieve the same result by asking people directly about their types and tell them directly what to do. The argument is quite obvious: Assume we have the equilibrium of the communication game. A mediator can ask the players to reveal their type and use the function r(.) to compute the reports that would have been sent under the equilibrium σ. Similarily he can use the function $\delta(.)$ to compute the actions that would have been chosen given the types were reported like in the equilibrium of the communication game and messages sent to them via the communication system. Since a mediator can do this he can tell directly to the players what actions to take according to his recommendation. Since we start from an equilibrium no player has an incentive to lie given all the others tell the truth or to disobey given all the others obey.

What is to my understanding the more interesting point of this result is the kind of construction how one can include communication possibilities into game theory in a practical way. Starting from a very general mediated communication game one shows by a construction like the revelation principle that the set of incentive compatible mediation plans for a general game with communication is essentially equivalent to the correlated equilibria of a game with communication characterized by an equivalent direct mechanism. It shows therefore an approach to include communication via the solution concept that depends only on the normal form thereby generating games that are in principle easy to analyze, because the set of incentive compatible direct mechanisms is mathematically nice in the sense that it is convex and compact and can be characterized by a finite set of linear inequalities. In this sense the revelation principle is a very powerful result.

If we construct the incentive compatible mechanism as described in the proof, we get exactly the set of correlated equilibria of the Bayesian game with communication with a simple communication structure.

3.3 The Principal agent model as a Bayesian Game with communication

Note that the setup described above is completely general so that all Principal-Agent models can be with a slight relabelling be subsumed under this class. In the framework developed above Principal-Agent models are a special form of games with communication. One player called *the principal* is in a position to be the first mover. Like in an ultimatum bargaining game he is in a position to make a take-it-or-leave-it offer about a particular mediation plan or mechanism to the other players. He can dictate the *rules of communication.*

The other players called *the agents* must react to the rules set by the principal and are willing to accept any Nash-equilibrium of the game the principal proposes. In a quite machiavellistic manner the problem to be considered is what the principal should do to squeeze the hell out of the agents or to put it in a more abstract way, to maximize his expected utility. So for example in the most prominent application of principal-agent models analyzing issues that are connected to the separation of ownership and control in modern firms, the owner (principal) wants to establish some coordination mechanism (the firm) to optimally exploit the managers and the employees (the agents).

Since now one player is in the position to design the communication system we have to add some minor changes to the setup from above. Reserve label 1 for the player called the principal. His problem is to coordinate his decision c_1 and the decisions and reports of all the others to maximize his expected utility. Since we are considering a principal agent problem the decicion of the principal may now also depend on the reports of the agents. Let μ denote the probability that the principal will choose a c_1 in C_1 and agents will receive messages m in M given the agents' plan to report r in R. So R,M together with μ describe the communication system. When we calculate the expected utilities we have now to take expectations of course with respect to the principal's decision as well. So the expected utility of a communication strategy for agent i becomes:

$$u_i((r_j(.), \delta_j(.))_{j \in N}) = \sum_{t \in T} \sum_{c_1 \in C_1} \sum_{m \in M} p(t)\mu(m(.) \mid r(.))u_i(\delta(m,t),t)$$

Now the problem is completely analogous to the setup described above. Given the general mechanism described above, we find a direct incentive compatible mechanism that maximizes the principals expected utility on a constraint set that has the nice form of a convex polyhedron so that we can state the principals' problem as a convex programming problem. By the revelation principle we know that this procedure will give us all implementable outcomes for the principal.

4 An example

In this section I want to present an example of pure adverse selection without moral hazard that builds on an example in David Kreps book "*A Course in Microeconomic Theory.*" I take this example because it gives a nice illustration of the quite abstract theory described above and at the same time points out a problem of arguing with the revelation principle. The revelation principle gives only a necessary condition for implementation. It says that *given* we have a Bayesian equilibrium of a game with communication there exists a direct incentive compatible mechanism in which every type of every player gets the same expected utility as in the game with communication. There is, however, nothing that guarantees that the players will choose the truthful equilibrium if other untruthful equilibria exist. These equilibria will however in general not be in the interest of the principal. So it might be possible that his preferred equilibrium of the game might not be implementable if the agents can somehow coordinate to play the untruthful one. As an extension to the example given in Kreps who points out this problem, Raffaelo Repullo has shown that it might very well be possible that in spite of the fact that the direct mechanism is not able to implement the truthtelling equilibrium, if the agents can coordinate on the untruthful one, there can exist an indirect mechanism that implements the truthtelling equilibrium. So we can get rid of an equilibrium by going from the direct to the indirect mechanism. This can be shown for instance for the Kreps example. Repullo gives a similar example which is however somewhat hermetically presented.[22].

Since the idea of an incentive contract is that the principal can overcome the information problem optimally by proposing a properly constructed game to the agents and the analysis predicts that they are willing to accept every Nash-equilibrium of this game, we have of course all the problems that we have with Nash equilibria in mechanism design problems as well. Since Nash equilibria are in general not unique we should expect to get several Nash equilibria as the rule. Given this multiplicity of equilibria we have the problem to decide which equilibrium seems more plausible to us to predict as the "solution" to the game. So we should be aware of the problem that a direct mechanism can generate several Nash equilibria among which the truthtelling equilibrium is not the most compelling one to choose as the solution. There might, however, be an indirect mechanism that does the job quite well.

Imagine a public authority who has to procure two units of some good, say liedetectors, for the municipal police. There are two suppliers in the market each

[22] See Repullo 1986.

one with production costs unknown to the authority. Since the authority has to pay whatever the supplier claims to be the price there is an obvious conflict of interest between the parties. Both suppliers have an incentive to claim as their costs as much as possible to justify their prices. Lets assume the authority wants to use taxrevenues in a parsimonious way and therefore would like to buy the detectors from the cheapest supplier. Since it hasn't got the detectors yet to decide whose claims are true the authority has to think of a mechanism that makes truthtelling a matter of self-interest to the parties.[23]

To advise the authority we can employ the theory developed above to find a smart way to get a truthful report. Note first that we have a problem of pure adverse selection. There is no imperfect information about the parties decision domains. All the parties can do is to report their costs the rest is decided by the authority.

Let each agent's type space be given by $T_i = \{1, 2\}$, denoting the privatly known costs per unit supplied. Let us assume that types are independent and that prior believes are given by $p_i(t_{-i} \mid t_i) = 0.5$, respectively. The payoffs for the agents are given by the function: $u_i(c_1 \mid t_i) = x_i(p_i - t_i)$

The governments' objective is to choose a fourtuple of prices and quantities of the following kind:

$$C_1 = \{x^1_{i,j}, x^2_{i,j}, p^1_{i,j}, p^2_{i,j} \in R^4 \mid x^1_{i,j} + x^2_{i,j} = 2\}$$

By the revelation principle we know that if there is an equilibrium of the communication game there is an equilibrium induced by a direct incentive compatible mechanism in which the players get the same expected utility so we can write this down as an optimization problem under a set of linear constraints.

$$\min_{c_1} \sum_{i,j \in \{1,2\} \times \{1,2\}} 0.25(p^1_{ij} x^2_{ij} + p^2_{ij} x^2_{ij})$$

[23]Formally this situation differs slightly from a Bayesian Game in that we have a situation where a set of outcomes or social options, denoted as D, is *jointly feasible* for the players together. So instead of the strategy sets that define a Bayesian game we have now a nonempty set of *social outcomes*. Formally this is therefore not a Bayesian game but a *Bayesian collective Choice Problem* (see Myerson [1991]) However a Bayesian collective choice problem subsumes a Bayesian game if we can write it in the form that there exists a mapping g from the cartesian product of the individual strategy sets C_i into D such that $u_i(c, t) = u_i(g(c), t) \quad \forall i \in N, \quad \forall c \in C, \quad \forall t \in T$ If μ is an incentive compatible mediation plan for the Bayesian game that is subsumed by the Bayesian collective choice problem then μ is incentive compatible for the Bayesian collective choice problem as well. In this sense a construction like the revelation principle can be applied to Bayesian collective choice problems by the arguments explained in this paper.

s.t.

$$\begin{aligned}
0.5x^1_{11}(p^1_{11}-1)+0.5x^1_{12}(p^1_{12}-1) &\geq 0\\
0.5x^1_{21}(p^1_{21}-2)+0.5x^1_{22}(p^1_{22}-2) &\geq 0\\
0.5x^2_{11}(p^2_{11}-1)+0.5x^2_{12}(p^2_{12}-1) &\geq 0\\
0.5x^2_{21}(p^1_{21}-2)+0.5x^2_{22}(p^2_{22}-2) &\geq 0
\end{aligned}$$

These constraints describe the *individual rationality* or[24] *participation constraints.* If we think of the extensive form of the game this expresses the idea of the principal moving first making the agents a take it or leave it offer. The mechanism he proposes must give the agents at least the value of their outside option. This requirement is expressed in the above constraints.

Second the *incentive constraints* have to be fulfilled to implement truth telling as an equilibrium. The agents have to get at least as much expected utility from truth telling as from lying.

$$\begin{aligned}
0.5x^1_{11}(p^1_{11}-1)+0.5x^1_{12}(p^1_{12}-1) &\geq 0.5x^1_{21}(p^1_{21}-1)+0.5x^1_{22}(p^1_{22}-1)\\
0.5x^1_{21}(p^1_{21}-2)+0.5x^1_{22}(p^1_{22}-2) &\geq 0.5x^1_{11}(p^1_{21}-2)+0.5x^1_{12}(p^1_{22}-2)\\
0.5x^2_{11}(p^2_{11}-1)+0.5x^2_{12}(p^2_{12}-1) &\geq 0.5x^2_{21}(p^2_{21}-1)+0.5x^2_{22}(p^2_{22}-1)\\
0.5x^2_{21}(p^2_{21}-2)+0.5x^2_{22}(p^2_{22}-2) &\geq 0.5x^2_{11}(p^2_{21}-2)+0.5x^2_{12}(p^2_{22}-2)
\end{aligned}$$

These inequalities characterize the relevant incentive constraints

Furthermore I impose the constraints that all prices and quantities shall be nonnegative and that the quantities add up to 2 in every combination of types. This is a standard optimization problem, which can be solved by the appropriate techniques.

A solution to the problem is the following optimal mechanism.[25]

	t_1=1				t_2=2			
t_2=1	(1	1	1.6	1.6)	(2	0	1.2	0)
t_2=2	(0	2	0	1.2)	(1	1	2	2)

[24]If this notation seems confusing take the following readers guide: x^1_{12} is to read as quantity x demanded from the player whose true type is 1 (superscript) given he reports to be type one and the other player reports to be type two.(subscript)

[25]To calculate this solution it is a good idea to use a computer because this simple problem is already somewhat huge to solve with pencil and paper

This mechanism says for example: If both players report costs 1 then the authority commits itself to demand one detector from firm one and one detector from firm two and pays both firms price 1.6. This mechanism induces a game of incomplete information among the agents, which has the following normal form as one can check by some easy but tedious calculation.

	C_2			
C_1	(1,2)	(1,1)	(2,1)	(2,2)
(1,2)	(.25,.25)	(.3,-.25)	(.25,-.25)	(.2,.25)
(1,1)	(-.25,.3)	(.1,.1)	(-.25,-.2)	(-.6,0)
(2,1)	(-.25,.25)	(-.2,-.25)	(-.25,-.25)	(-.3,.25)
(2,2)	(.25,.2)	(0;-.6)	(.25,-.25)	(.5,.5)

As one can easily check now this game has two Nash equilibria. One equilibrium at the upperleftcorner, where both of the players reveal their type truthfully. This should be no surprise because we constructed the mechanism such that truthtelling is a Bayesian equilibrium for the game we are interested in. Note however that there is a second Bayesian equilibrium at the lower right corner in which both players lie about their true costs. Compared to the other equilibrium this equilibrium is strict, i.e. it is the only best reply to itself. The truthtelling equilibrium however is not strict. Furthermore the former equilibrium pareto dominates the truthtelling one. If the players would ever have a chance to somehow coordinate they would be better off in the nontruthful equilibrium. So one can argue that if the players play the lying equilibrium the government can't implement the truthful equilibrium by the direct mechanism. This is a sad situation for the authority because the most attractive Nash equilibrium for the agents is at the same time the most expensive solution for the government.

But there is still some hope to find out the truth by game theoretic reasoning. Instead of staring at the direct mechanism we allow the players to communicate via an indirect mechanism of the following kind. We let their report space now consist of the statements $R_i = \{1, 2, 3\}$ Now we construct the following mechanism:

	r_2=1	r_2=2	r_2=3
r_1=1	(1 1 1.6 1.6)	(2 0 1.2 0)	(1 1 3 0)
r_1=2	(0 2 0 1.2)	(1 1 2 2)	(1 1 0 3)
r_1=3	(1 1 0 3)	(1 1 3 0)	(1 1 0 0)

As before it is possible to write down the game of incomplete information induced by this mechanism, which is easy in principle but somewhat more tedious than before. [26] Checking for the Nash equilibrium yields that the truth telling equilibrium now is in fact the unique pure Nash equilibrium of the game and will therefore implement the equilibrium the principal likes most. It is also the one that yields minimal costs for the procured good. So it is not quite correct to state that without loss of generality we can use the revelation principle to check for implementable outcomes. In this example the indirect mechanism does but the direct mechanism does not implement the truth telling equilibrium at least if we select the lying equilibrium on the grounds that it is strict and paretodominant.

5 Conclusions

In this paper I have tried to make clear that all interesting applications of asymmetric information problems can be treated in a unified way by using the language of game theory. I have argued that an appropriate way to take communication possibilities that are inherent in principal agent problems into account is via the solution concept. By this approach we can identify the solutions of incentive compatible mediation plans by the correlated equilibria of the normal form game from which a particular game with communication is induced. The set of correlated equilibria has a nice mathematical structure because it convex and compact and can be described by a finite set of linear inequalities. The game theoretic approach at the same time makes clear where we have to be careful if we use the revelation principle. If an incentive compatible mediation plan is implemented as a Nash equilibrium the problem of multiple equilibria can arise and in particular the agents might not play the equilibrium the principal wants to implement. Since the revelation principle gives only a necessary condition for implementation it is not correct to state that we can use it without loss of generality. The main result of the theoretical arguments is that we can make life much simpler as one might think from the outset. But this simplicity is a main result of the theory.

[26] Observe that player one has now three options, whereas player two can respond by three options not knowing the type of the other player and vice versa. We therefore have to consider nine cases for each of the players, which gives us a nine by nine matrix of the normal form. A sheet of skratch paper with this normal form is available from me. I had to do it to check if Klaus told us the truth or not, and since we weren't smart enough to put him into an incentive compatible mechanism I had to check it all by my own. As it turned out he was lying a little bit. The truth telling equilibrium is not strict, but it is the only one in pure strategies. There are also some equilibria in mixed strategies.

6 References

Antle, R.[1982] *The Auditor as an economic agent*, Journal of Accounting Research, vol.20, no. 2, 503-527

Aumann, R. [1974] *Subjectivity and Correlation in Randomized Strategies*, Journal of Mathematical Economics, 1, p 67-96

Aumann, R. [1976] *Agreeing to disagree*, Annals of Statistics,4, 1236-39

Aumann, R. [1987] *Correlated Equilibrium as an Expression of Bayesian Rationality*, Econometrica 55, Vol.1, pp 1-18

Dasgupta, P., Hammond, Maskin [1979] *The implementation of social choice rules. Some results on incentive compatibility.*, Review of Economic Studies 46, 185-216

Forges, F. [1986] *An approach to communication equilibria.*, Econometrica, vol 54, No 6, 1375-85

Gibbard, [1973] *Manipulation of Voting Schemes. A general result*, Econometrica 41, 587-602

Harsanyi, J.C. [1967/68] *Games with Incomplete Information Played by Bayesian Players*, Management Science 14, 159-182, 159-182, 320-334, 486-502

Harsanyi, J.C., Selten, R. [1988] *A General Theory of Equilibrium Selection in Games*, MIT Press

Kreps, D.[1990]*A Course in Microeconomic Theory*, Harvester Wheatsheaf

Myerson, R.B. [1982]*Optimal Coordination Mechanisms in Generalized Principal Agent Problems*, Journal of Mathematical Economics, 10, 67-81

Myerson, R.B. [1991] *Game theory. Analysis of Conflict*, Harvard University Press

Repullo, R. [1986] *On the Revelation Principle under Complete and Incomplete Information*, in **Binmore, Dasgupta** (eds.): *Economic Organizations as games*, Basil Blackwell, Oxford

Spence, M. [1973] *Job Market Signalling*, Quarterly Journal of Economics 87, 355-374

Varian, H. [1992] *Microeconomic Theory*, Norton