Commitment and Extortion*

Paul Harrenstein
University of Munich
80538 Munich, Germany
harrenst@tcs.ifi.lmu.de

Felix Brandt
University of Munich
80538 Munich, Germany
brandtf@tcs.ifi.lmu.de

Felix Fischer
University of Munich
80538 Munich, Germany
fischerf@tcs.ifi.lmu.de

ABSTRACT

Making commitments, *e.g.*, through promises and threats, enables a player to exploit the strengths of his own strategic position as well as the weaknesses of that of his opponents. Which commitments a player can make with credibility depends on the circumstances. In some, a player can only commit to the performance of an action, in others, he can commit himself *conditionally* on the actions of the other players. Some situations even allow for commitments on commitments or for commitments to randomized actions. We explore the formal properties of these types of (conditional) commitment and their interrelationships. So as to preclude inconsistencies among conditional commitments, we assume an order in which the players make their commitments. Central to our analyses is the notion of an *extortion*, which we define, for a given order of the players, as a profile that contains, for each player, an optimal commitment given the commitments of the players that committed earlier. On this basis, we investigate for different commitment types whether it is advantageous to commit earlier rather than later, and how the outcomes obtained through extortions relate to backward induction and Pareto efficiency.

General Terms

Economics, Theory

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*

Keywords

Multiagent Systems, Game Theory, Commitment, Extortion

1. INTRODUCTION

On one view, the least one may expect of game theory is that it provides an answer to the question which actions maximize an agent's expected utility in situations of interactive decision making. A slightly divergent view is expounded by Schelling when he states that "strategy [. . .] is not concerned with the efficient *application* of force but with the *exploitation of potential force*" [9, page 5]. From this perspective, the formal model of a game in strategic form only outlines the strategic features of an interactive situation. Apart from merely choosing and performing an action from a set of actions, there may also be other courses open to an agent. *E.g.*, the strategic lie of the land may be such that a promise, a threat, or a combination of both would be more conductive to his ends.

The potency of a promise, however, essentially depends on the extent the promisee can be convinced of the promiser's resolve to see to its fulfillment. Likewise, a threat only succeeds in deterring an agent if the latter can be made to believe that the threatener is bound to execute the threat, should it be ignored. In this sense, promises and threats essentially involve a *commitment* on the part of the one who makes them, thus purposely restricting his freedom of choice. Promises and threats epitomize one of the fundamental and at first sight perhaps most surprising phenomena in game theory: it may occur that a player can improve his strategic position *by limiting his own freedom of action.* By *commitments* we will understand such limitations of one's action space. Action itself could be seen as the ultimate commitment. Performing a particular action means doing so to the exclusion of all other actions.

Commitments come in different forms and it may depend on the circumstances which ones can and which ones cannot credibly be made. Besides simply committing to the performance of an action, an agent might make his commitment *conditional* on the actions of other agents, as, *e.g.*, the kidnapper does, when he promises to set free a hostage on receiving a ransom, while threatening to cut off another toe, otherwise. Some situations even allow for commitments on commitments or for commitments to randomized actions.

By focusing on the selection of actions rather than on commitments, it might seem that the conception of game theory as mere interactive decision theory is too narrow. In this respect, Schelling's view might seem to evince a more comprehensive understanding of what game theory tries to accomplish. One might object, that commitments could be seen as the actions of a larger game. In reply to this criticism Schelling remarks:

> While it is instructive and intellectually satisfying to see how such tactics as threats, commitments, and promises can be absorbed in an enlarged, abstract "supergame" (game in "normal form"), it should be emphasized that we cannot learn anything about those tactics by studying games that are already in normal form. [. . .] What we want is a theory that systematizes the study of the various universal ingredients that make up the move-structure of games; too abstract a model will miss them. [9, pp. 156-7]

*This material is based upon work supported by the Deutsche Forschungsgemeinschaft under grant BR 2312/3-1.

AAMAS'07 May 14–18 2007, Honolulu, Hawai'i, USA.

Our concern is with these commitment tactics, be it that our analysis is confined to situations in which the players can commit in a given order and where we assume the commitments the players can make are given. Despite Schelling's warning for too abstract a framework, our approach will be based on the formal notion of an *extortion*, which we will propose in Section 4 as a uniform tactic for a comprehensive class of situations in which commitments can be made sequentially. On this basis we tackle such issues as the usefulness of certain types of commitment in different situations (strategic games) or whether it is better to commit early rather than late. We also provide a framework for the assessment of more general game theoretic matters like the relationship of extortions to backward induction or Pareto efficiency.

Insight into these matters has proved itself invaluable for a proper understanding of diplomatic policy during the Cold War. Nowadays, we believe, these issues are equally significant for applications and developments in such fields as multiagent systems, distributed computing and electronic markets. For example, commitments have been argued to be of importance for *interacting software agents* as well as for *mechanism design*. In the former setting, the inability to re-program a software agent on the fly can be seen as a commitment to its specification and thus exploited to strengthen its strategic position in a multiagent setting. A mechanism, on the other hand, could be seen as a set of commitments that steers the players' behavior in a certain desired way (see, *e.g.*, [2]).

Our analysis is conceptually similar to that of *Stackelberg* or *leadership games* [15], which have been extensively studied in the economic literature (*cf.*, [16]). These games analyze situations in which a *leader* commits to a pure or mixed strategy, and a number of *followers*, who then act simultaneously. Our approach, however, differs in that it is assumed that the players all move in a particular order—first, second, third and so on—and that it is specifically aimed at incorporating a wide range of possible commitments, in particular conditional commitments.

After briefly discussing related work in Section 2, we present the formal game theoretic framework, in which we define the notions of a *commitment type* as well as *conditional* and *unconditional commitments* (Section 3). In Section 4 we propose the generic concept of an *extortion*, which for each commitment type captures the idea of an optimal commitment profile. We point out an equivalence between extortions and backward induction solutions, and investigate whether it is advantageous to commit earlier rather than later and how the outcomes obtained through extortions relate to Pareto efficiency. Section 5 briefly reviews some other commitment types, such as *inductive*, *mixed* and *mixed conditional commitments*. The paper concludes with an overview of the results and an outlook for future research in Section 6.

2. RELATED WORK

Commitment is a central concept in game theory. The possibility to make commitments distinguishes cooperative from non-cooperative game theory [4, 6]. Leadership games, as mentioned in the introduction, analyze commitments to pure or mixed strategies in what is essentially a two-player setting [15, 16]. Informally, Schelling [9] has emphasized the importance of promises, threats and the like for a proper understanding of social interaction. On a more formal level, threats have also figured in bargaining theory. Nash's *threat game* [5] and Harsanyi's *rational threats* [3] are two important early examples. Also, commitments have played a significant role in the theory of *equilibrium selection* (see, *e.g.*, [13].

Over the last few years, game theory has become almost indispensable as a research tool for computer science and (multi)agent research. Commitments have by no means gone unnoticed (see, *e.g.*, [1, 11]). Recently, also the *strategic* aspects of commitments have attracted the attention of computer scientists. Thus, Conitzer and Sandholm [2] have studied the computational complexity of computing the optimal strategy to commit to in normal form and Bayesian games. Sandholm and Lesser [8] employ levelled commitments for the design of multiagent systems in which contractual agreements are not fully binding. Another connection between commitments and computer science has been pointed out by Samet [7] and Tennenholtz [12]. Their point of departure is the observation that programs can be used to formulate commitments that are conditional on the programs of other systems.

$$\begin{bmatrix} (1,3) & (3,2) \\ (0,0) & (2,1) \end{bmatrix}$$

Figure 1: Committing to a dominated strategy can be advantageous.

Our approach is similar to the Stackleberg setting in that we assume an order in which the players commit. We, however, consider a number of different commitment types, among which conditional commitments, and propose a generic solution concept.

3. COMMITMENTS

By committing, an agent can improve his strategic position. It may even be advantageous to commit to a strategy that is *strongly dominated*, *i.e.*, one for which there is another strategy that yields a better payoff no matter how the other agents act. Consider for example the 2×2 game in Figure 1, in which one player, *Row*, chooses rows and another, *Col*, chooses columns. The entries in the matrix indicate the payoffs to *Row* and *Col*, respectively. Then, top-left is the solution obtained by iterative elimination of strongly dominated strategies: for *Row*, playing top is always better than playing bottom, and assuming that *Row* will therefore never play bottom, left is always better than right for *Col*. However, if *Row* succeeds in convincing *Col* of his commitment to play bottom, the latter had better choose the right column. Thus, *Row* attains a payoff of two instead of one. Along a similar line of reasoning, however, *Col* would wish to commit to the left column, as convincing *Row* of this commitment guarantees him the most desirable outcome. If, on the other hand, both players actually commit themselves in this way but *without* convincing the other party of their having done so, the game ends in misery for both.

Important types of commitments, however, cannot simply be analyzed as unconditional commitments to actions. The essence of a threat, for example, is deterrence. If successful, it is *not* carried out. (This is also the reason why the credibility of a threat is not necessarily undermined if its putting into effect means that the threatener is also harmed.) By contrast, promises are made to entice and, as such, meant to be fulfilled. Thus, both threats and promises would be strategically void if they were unconditional.

Figure 2 shows an example, in which *Col* can guarantee himself a payoff of three by threatening to choose the right column if *Row* chooses top. (This will suffice to deter *Row*, and there is no need for an additional promise on the part of *Col*.) He cannot do so by merely committing unconditionally, and neither can *Row* if he were to commit first.

In the case of *conditional* commitments, however, a particular kind of inconsistency can arise. It is not in general the case that any two commitments can both be credible. In a 2×2 game, it could occur that *Row* commits conditionally on playing top if the

$$\begin{bmatrix} (2,2) & (0,0) \\ (1,3) & (3,1) \end{bmatrix}$$

Figure 2: The column player *Col* can guarantee himself a payoff of three by threatening to play right if the row player *Row* plays top.

Col plays left, and bottom, otherwise. If now, *Col* simultaneously were able to commit to the conditional strategy to play right if *Row* plays top, and left, otherwise, there is no strategy profile that can be played without one of the players' bluff being called.

To get around this problem, one can write down conditional commitments in the form of rules and define appropriate fixed point constructions, as suggested by Samet [7] and Tennenholtz [12]. Since checking the semantic equivalence of two commitments (or commitment conditions) is undecidable in general, Tennenholtz bases his definition of *program equilibrium* on syntactic equivalence. We, by contrast, try to steer clear from fixed point constructions by assuming that the players make their commitment in a particular order. Each player can then make his commitments dependent on the actions of the players to commit after him, but not on the commitments of the players that committed before. On the issue how this order comes about we do not here enter. Rather, we assume it to be determined by the circumstances, which may force or permit some players to commit earlier and others later. We will find that it is not always beneficial to commit earlier than later or *vice versa*.

Another point to heed is that we only consider the case in which the commitments are considered *absolutely binding*. We do not take into account commitments that can be violated. Intuitively, this could be understood as that the possibility of violation fatally undermines the credibility of the commitment. We also assume commitments to be *complete*, in the sense that they fully lay down a player's behavior in all foreseeable circumstances. These assumptions imply that the outcome of the game is entirely determined by the commitments the players make. Although these might be implausible assumptions for some situations, we had better study the idealized case first, before tackling the complications of the more general case. To make these concepts formally precise, we first have to fix some notation.

3.1 Strategic Games

A *strategic game* is a tuple $(N, (A_i)_{i\in N}, (u_i)_{i\in N})$, where $N = \{1, \ldots, n\}$ is a finite set of players, A_i is a set of actions available to player i and u_i a real-valued utility function for player i on the set of *(pure) strategy profiles* $S = A_1 \times \cdots \times A_n$. We call a game *finite* if for all players i the action set A_i is finite. A *mixed strategy* σ_i for a player i is a probability distribution over A_i. We write Σ_i for the set of mixed strategies available to player i, and $\Sigma = \Sigma_1 \times \cdots \times \Sigma_n$ for the set of *mixed strategy profiles*. We further have $\sigma(a)$ and $\sigma_i(a)$ denote the probability of action a in mixed strategy profile σ or mixed strategy σ_i, respectively. In settings involving expected utility, we will generally assume that utility functions represent von Neumann-Morgenstern preferences. For a player i and (mixed) strategy profiles σ and τ we write $\sigma \preccurlyeq_i \tau$ if $u_i(\sigma) \leqslant u_i(\tau)$.

3.2 Conditional Commitments

Relative to a strategic game $(N, (A_i)_{i\in N}, (u_i)_{i\in N})$ and an ordering $\pi = (\pi_1, \ldots, \pi_n)$ of the players, we define the set F_{π_i} of *(pure) conditional commitments* of a player π_i as the set of functions from $A_{\pi_1} \times \cdots \times A_{\pi_{i-1}}$ to A_{π_i}. For π_1 we have the set of conditional commitments coincide with A_{π_1}. By a *conditional commitment profile* f we understand any combination of conditional commitments in $F_{\pi_1} \times \cdots \times F_{\pi_n}$.

Intuitively, π reflects the sequential order in which the players can make their commitments, with π_n committing first, π_{n-1} second, and so on. Each player can condition his action on the actions of all players that are to commit *after* him. In this manner, each conditional commitment profile f can be seen to determine a unique strategy profile, denoted by f', which will be played if all players stick to their conditional commitments. More formally, the strategy profile $f' = (f'_{\pi_1}, \ldots, f'_{\pi_n})$ for a conditional commitment profile f is defined inductively as

$$\begin{aligned} f'_{\pi_1} &=_{df.} f_{\pi_1}, \\ f'_{\pi_{i+1}} &=_{df.} f_{\pi_{i+1}}(f'_{\pi_1}, \ldots, f'_{\pi_i}). \end{aligned}$$

The sequence $f'_{\pi_1}, (f'_{\pi_1}, f'_{\pi_2}), \ldots, (f'_{\pi_1}, \ldots, f'_{\pi_n})$ will be called the *path of* f. *E.g.*, in the two-player game of Figure 2 and given the order (Row, Col), *Row* has two conditional commitments, top and bottom, which we will henceforth denote t and b. *Col*, on the other hand, has four conditional commitments, corresponding to the different functions mapping strategies of *Row* to those of *Col*. If we consider a conditional commitment f for *Col* such that $f(t) = l$ and $f(b) = r$, then (t, f) is a conditional commitment profile and $(t, f)' = (t, f(t)) = (t, l)$.

There is a natural way in which a strategic game G together with an ordering $(\pi_1, \ldots, \pi_n)$ of the players can be interpreted as an *extensive form game with perfect information* (see, *e.g.*, [4, 6])[1], in which π_1 chooses his action first, π_2 second, and so on. Observe that under this assumption the *strategies* in the extensive form game and the *conditional commitments* in the strategic game G with ordering π are mathematically the same objects. Applying *backward induction* to the extensive form game yields *subgame perfect equilibria*, which arguably provide appropriate solutions in this setting. From the perspective of conditional commitments, however, players move in reverse order. We will argue that under this interpretation other strategy profiles should be singled out as appropriate.

To illustrate this point, consider once more the game in Figure 2 and observe that neither player can improve on the outcome obtained via iterated strong dominance by committing unconditionally to some strategy. Situations like this, in which players can make unconditional commitments *in a fixed order*, can fruitfully be analyzed as extensive form games, and the most lucrative unconditional commitment can be found through *backward induction*. Figure 3 shows the extensive form associated with the game of Figure 2. The strategies available to the row player are the same as in the strategic form: choosing the top or the bottom row. The strategies for the column player in the extensive game are given by the four functions that map strategies of the row player in the strategic game to one of his own. Transforming this extensive form back into a strategic game (see Figure 4), we find that there exists a second equilibrium besides the one found by means of backward induction. This equilibrium with outcome $(1, 3)$, indicated by the thick lines in Figure 3, has been argued to be unacceptable in the sequential game as it would involve an *incredible threat* by *Col*: once *Row* has played top, *Col* finds himself confronted with a *fait accompli*. He had better make the best of a bad bargain and opt for the left column after all. This is in essence the line of thought Selten followed in his famous argument for *subgame perfect equilibria* [10]. If, however, the strategies of *Col* in the extensive form are thought of as his *conditional* commitments he can make in case

[1] For a formal definition of a game in extensive form, the reader consult one of the standard textbooks, such as [4] or [6]. In this paper all formal definitions are based on strategic games and orderings of the players only.

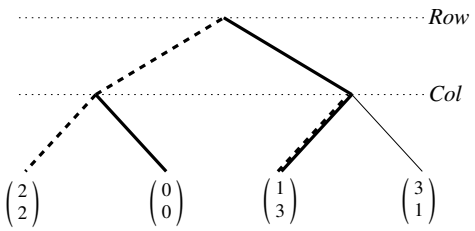

Figure 3: Extensive form obtained from the strategic game of Figure 2 when the row player chooses an action first. The backward induction solution is indicated by dashed lines, the conditional commitment solution by solid ones. (The horizontal dotted lines do *not* indicate information sets, but merely indicate which players are to move when.)

he moves first, the situation is radically different. Thus we also assume that it is possible for *Col* to make credible the threat to choose the right column if *Row* were to play top, so as to ensure the latter is always better off to play the bottom row. If *Col* can make a conditional commitment of playing the right column if *Row* chooses top, and the left column otherwise, this leaves *Row* with the easy choice between a payoff of zero or one, and *Col* may expect a payoff of three.

This line of reasoning can be generalized to yield an algorithm for finding optimal conditional commitments for general two-player games:

1. Find a strategy profile $s = (s_{\pi_1}, s_{\pi_2})$ with maximum payoff to player π_2, and set $f_{\pi_1} = s_{\pi_1}$ and $f_{\pi_2}(s_{\pi_1}) = s_{\pi_2}$.
2. For each $t_{\pi_1} \in A_{\pi_1}$ with $t_{\pi_1} \neq s_{\pi_1}$, find a strategy $t_{\pi_2} \in A_{\pi_2}$ that minimizes $u_{\pi_1}(t_{\pi_1}, t_{\pi_2})$, and set $f_{\pi_2}(t_{\pi_1}) = t_{\pi_2}$.
3. If $u_{\pi_1}(t_{\pi_1}, f_{\pi_2}(t_{\pi_1})) \leqslant u_{\pi_1}(s_{\pi_1}, s_{\pi_2})$ for all $t_{\pi_1} \neq s_{\pi_1}$, return f.
4. Otherwise, find the strategy profile (s'_{π_1}, s'_{π_2}) with the highest payoff to π_2 among the ones that have not yet been considered. Set $f_{\pi_1} = s'_{\pi_1}$ and $f_{\pi_2}(s'_{\pi_1}) = s'_{\pi_2}$, and continue with Step 2.

Generalizing the idea underlying this algorithm, we present in Section 4 the concept of an *extortion*, which applies to games with any number of players. For any order of the players an extortion contains, for each player, an optimal commitment given the commitments of the players that committed earlier.

3.3 Commitment Types

So far, we have distinguished between conditional and unconditional commitments. If made sequentially, both of them determine a unique strategy profile in a given strategic game. This notion of sequential commitment allows for generalization and gives rise to the following definition of a *(sequential) commitment type*.

Definition 3.1. *(Sequential commitment type) A* (sequential) commitment type τ *associates with each strategic game G and each ordering π of its players, a tuple $(X_{\pi_1}, \ldots, X_{\pi_n}, \phi)$, where $X_{\pi_1}, \ldots, X_{\pi_n}$ are (abstract) sets of commitments and ϕ is a function mapping each profile in $X = X_{\pi_1} \times \cdots \times X_{\pi_n}$ to a (mixed) strategy profile of G. A commitment type $(X_{\pi_1}, \ldots, X_{\pi_n}, \phi)$ is* finite *whenever X_{π_i} is finite for each i with $1 \leqslant i \leqslant n$.*

Thus, the type of *unconditional* commitments associates with a game and an ordering π of its players the tuple $(S_{\pi_1}, \ldots, S_{\pi_n}, id)$,

$$\begin{bmatrix} \mathbf{(2,2)} & (2,2) & (0,0) & (0,0) \\ (1,3) & (3,1) & \mathbf{(1,3)} & (3,1) \end{bmatrix}$$

Figure 4: The strategic game corresponding to the extensive form of Figure 3

where *id* is the identity function. Similarly, $(F_{\pi_1}, \ldots, F_{\pi_n}, ')$ is the tuple associated with the same game by the type of *(pure) conditional* commitments.

4. EXTORTIONS

In the introduction, we argued informally how players could improve their position by conditionally committing. How well they can do, could be analyzed by means of an extensive game with the actions of each player being defined as the possible commitments he can make. Here, we introduce for each commitment type a corresponding notion of *extortion*, which is defined relative to a strategic game and an ordering of the players. Extortions are meant to capture the concept of a profile that contains, for each player, an optimal commitment given the commitments of the players that committed earlier. A complicating factor is that in finding a player's optimal commitment, one should not only take into account how such a commitment affects other players' actions, but also how it enables them to make their commitments.

Definition 4.1. *(Extortions) Let G be a strategic game, π an ordering of its players, and τ a commitment type. Let $\tau(G, \pi)$ be given by $(X_{\pi_1}, \ldots, X_{\pi_n}, \phi)$. A* τ-extortion *of order 0 is any commitment profile $x \in X_{\pi_1} \times \cdots \times X_{\pi_n}$. For $m > 0$, a commitment profile $x \in X_{\pi_1} \times \cdots \times X_{\pi_n}$ is a* τ-extortion of order m in G given π *if x is an τ-extortion of order $m-1$ with*

$$\phi(y_{\pi_1}, \ldots, y_{\pi_m}, x_{\pi_{m+1}}, \ldots, x_{\pi_n}) \preccurlyeq_{\pi_m} \phi(x_{\pi_1}, \ldots, x_{\pi_m}, x_{\pi_{m+1}}, \ldots, x_{\pi_n})$$

for all commitment profiles g in X with $(y_{\pi_1}, \ldots, y_{\pi_m}, x_{\pi_{m+1}}, \ldots, x_{\pi_n})$ a τ-extortion of order $m-1$. A τ-extortion *is a commitment profile that is a τ-extortion of order m for all m with $0 \leqslant m \leqslant n$. Furthermore, we say that a (mixed) strategy profile σ is* τ-extortionable *if there is some τ-extortion x with $\phi(x) = s$.*

Thus, an extortion of order 1 is a commitment profile in which player π_1, makes a commitment that maximizes his payoff, *given fixed commitments of the other players*. An extortion of order m is an extortion of order $m-1$ that maximizes player m's payoff, *given fixed commitments of the players π_{m+1} through π_n*.

For the type of *conditional commitments* we have that any conditional commitment profile f is an extortion of order 0 and an extortion of an order m greater than 0 is any extortion of order $m-1$ for which:

$$(g_{\pi_1}, \ldots, g_{\pi_m}, f_{\pi_{m+1}}, \ldots, f_{\pi_n})' \preccurlyeq_{\pi_m} (f_{\pi_1}, \ldots, f_{\pi_m}, f_{\pi_{m+1}}, \ldots, f_{\pi_n})',$$

for each conditional commitment profile g such that $(g_{\pi_1}, \ldots, g_{\pi_m}, f_{\pi_{m+1}}, \ldots, f_{\pi_n})$ an extortion of order $m-1$.

To illustrate the concept of an extortion for conditional commitments consider the three-player game in Figure 5 and assume

$$\begin{bmatrix} (1,4,0) & (1,4,0) \\ (3,3,2) & (0,0,2) \end{bmatrix} \qquad \begin{bmatrix} (4,1,1) & (4,0,0) \\ (3,3,2) & (0,0,2) \end{bmatrix}$$

Figure 5: A three-player strategic game

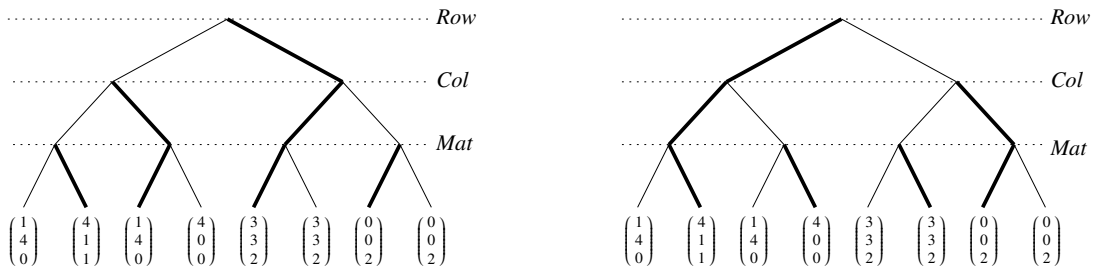

Figure 6: A conditional extortion f of order 1 (left) and an extortion g of order 3 (right).

(*Row*, *Col*, *Mat*) to be the order in which the players commit. Figure 6 depicts the possible conditional commitments of the players in extensive forms, with the left branch corresponding to *Row*'s strategy of playing the top row. Let f and g be the conditional commitment strategies indicated by the thick lines in the left and right figures respectively. Both f and g are extortions of order 1. In both f and g *Row* guarantees himself the higher payoff given the conditional commitments of *Mat* and *Col*. Only g, however, is also an extortion of order 2. To appreciate that f is not, consider the conditional commitment profile h in which *Row* chooses top and *Col* chooses right no matter how *Row* decides, *i.e.*, h is such that $h_{Row} = t$ and $h_{Col}(t) = h_{Col}(b) = r$. Then, $(h_{Row}, h_{Col}, f_{Mat})$ is also an extortion of order 1, but yields *Col* a higher payoff than f does. We leave it to the reader to check that, by contrast, g is an extortion of order 3, and therewith an extortion *per se*.

4.1 Promises and Threats

One way of understanding conditional extortions is by conceiving of them as combinations of precisely one promise and a number of threats. From the strategy profiles that can still be realized given the conditional commitments of players that have committed before him, a player tries to enforce the strategy profile that yields him as much payoff as possible. Hence, he chooses his commitment so as to render deviations from the path that leads to this strategy profile as unattractive as possible ('threats') and the desired strategy profile as appealing as possible ('promises') for the relevant players. If $(s_{\pi_1}, \ldots, s_{\pi_n})$ is such a desirable strategy profile for player π_i and f_{π_i} his conditional commitment, the value of $f_{\pi_i}(s_{\pi_1}, \ldots, s_{\pi_{i-1}})$ could be taken as his promise, whereas the values of f_{π_i} for all other $(t_{\pi_1}, \ldots, t_{\pi_{i-1}})$ could be seen as constituting his threats. The higher the payoff is to the other players in a strategy profile a player aims for, the easier it is for him to formulate an effective threat. However, making appropriate threats in this respect does not merely come down to minimizing the payoffs to players to commit later wherever possible. A player should also take into account the commitments, promises and threats the following players can make on the basis of his and his predecessors' commitments. This is what makes extortionate reasoning sometimes so complicated, especially in situations with more than two players.

For example, in the game of Figure 5, there is no conditional extortion that ensures *Mat* a payoff of two. To appreciate this, consider the possible commitments *Mat* can make in case *Row* plays top and *Col* plays left (*tl*) and in case *Row* plays top and *Col* plays right (*tr*). If *Mat* commits to the right matrix in both cases, he virtually promises *Row* a payoff of four, leaving himself with a payoff of at most one. Otherwise, he puts *Col* in a position to deter *Row* from choosing bottom by threatening to choose the right column if the latter does so. Again, *Mat* cannot expect a payoff higher than one. In short, no matter how *Mat* conditionally commits, he will either enable *Col* to threaten *Row* into playing top or fail to lure *Row* into playing the bottom row.

4.2 Benign Backward Induction

The solutions extortions provide can also be obtained by modeling the situation as an extensive form game and applying a backward inductive type of argument. The actions of the players in any such extensive form game are then given by their conditional commitments, which they then choose sequentially. For higher types of commitment, such as conditional commitments, such 'meta-games', however, grow exponentially in the number of strategies available to the players and are generally much larger than the original game. The correspondence between the backward induction solutions in the meta-game and the extortions of the original strategic game rather signifies that the concept of an extortion is defined properly. First we define the concept of *benign backward induction* in general relative to a game in *strategic form* together with an ordering of the players. Intuitively it reflects the idea that each player chooses for each possible combination of actions of his predecessors the action that yields the highest payoff, given that his successors do similarly. The concept is called *benign* backward induction, because it implies that a player, when indifferent between a number of actions, chooses the one that benefits his *predecessors* most. For an ordering π of the players, we have π^R denote its reversal $(\pi_n, \ldots, \pi_1)$.

Definition 4.2. *(Benign backward induction) Let G be a strategic game and π an ordering of its players. A* benign backward induction of order 0 *is any conditional commitment profile f subject to π. For $m > 0$, a conditional commitment strategy profile f is a benign backward induction (solution) of order m if f is a benign backward induction of order $m-1$ and*

$$(g_{\pi_n^R}, \ldots, g_{\pi_{m+1}^R}, g_{\pi_m^R}, \ldots, g_{\pi_1^R})' \preccurlyeq_{\pi_m^R} (g_{\pi_n^R}, \ldots, g_{\pi_{m+1}^R}, f_{\pi_m^R}, \ldots, f_{\pi_1^R})'$$

for any backward induction $(g_{\pi_n^R}, \ldots, g_{\pi_{m+1}^R}, g_{\pi_m^R}, \ldots, g_{\pi_1^R})$ of order $m-1$. A conditional commitment profile f is a benign backward induction *if it is a benign backward induction of order k for each k with $0 \leqslant k \leqslant n$.*

For games with a finite action set for each player, the following result follows straightforwardly from *Kuhn's Theorem* (*cf.* [6, p. 99]). In particular, this result holds if the players' actions are commitments of a finite type.

Fact 4.3. *For each finite game and each ordering of the players, benign backward inductions exist.*

For each game, each ordering of its players and each commitment type, we can define another game G^* with the the actions of each player i given by his τ-commitments X_i in G. The utility

of a *strategy profile* $(x_{\pi_1}, \ldots, x_{\pi_n})$ for a player i in G^* can then be equated to his utility of the strategy profile $\phi(x_{\pi_n}, \ldots, x_{\pi_1})$ in G. We now find that the extortions of G can be retrieved as the paths of the benign backward induction solutions of the game G^* for the ordering π^R of the players, provided that the commitment type is finite.

THEOREM 4.4. *Let $G = (N, (A_i)_{i\in N}, (u_i)_{i\in N})$ be a game and π an ordering of its players with which the finite commitment type τ associates the tuple $(X_{\pi_1}, \ldots, X_{\pi_n}, \phi)$. Let further $G^* = (N, (X_{\pi_i})_{i\in N}, (u^*_{\pi_i})_{i\in N})$, where $u^*_{\pi_i}(x_{\pi_n}, \ldots, x_{\pi_1}) = u_{\pi_i}(\phi(x_{\pi_1}, \ldots, x_{\pi_n}))$, for each τ-commitment profile $(x_{\pi_1}, \ldots, x_{\pi_n})$. Then, a π-commitment profile $(x_{\pi_1}, \ldots, x_{\pi_n})$ is a τ-extortion in G given π if and only if there is some benign backward induction f in G^* given π^R with $f' = (x_{\pi_n}, \ldots, x_{\pi_1})$.*

PROOF. Assume that f is a benign backward induction in G^* relative to π^R. Then, $f' = (x_{\pi_n}, \ldots, x_{\pi_1})$, for some commitment profile $(x_{\pi_1}, \ldots, x_{\pi_n})$ of G relative to π. We show by induction that $(x_{\pi_1}, \ldots, x_{\pi_n})$ is an extortion of order m, for all m with $0 \leqslant m \leqslant n$. For $m = 0$, the proof is trivial. For the induction step, consider an arbitrary commitment profile $(y_{\pi_1}, \ldots, y_{\pi_n})$ such that $(y_{\pi_1}, \ldots, y_{\pi_m}, x_{\pi_{m+1}}, \ldots, x_{\pi_n})$ is an extortion of order $m-1$. In virtue of the induction hypothesis, there is a benign backward induction g of order $m-1$ in G^* with $g' = (x_{\pi_n}, \ldots, x_{\pi_{m+1}}, y_{\pi_m}, \ldots, y_{\pi_1})$. As f is also a benign backward induction of order m:

$$(g_{\pi_n}, \ldots, g_{\pi_1})' \leqslant^*_{\pi_m} (g_{\pi_n}, \ldots, g_{\pi_{m+1}}, f_{\pi_m}, \ldots, f_{\pi_1})'.$$

Hence, $(x_{\pi_n}, \ldots, x_{\pi_{m+1}}, y_{\pi_m}, \ldots, y_{\pi_1}) \leqslant^*_{\pi_m} (x_{\pi_n}, \ldots, x_{\pi_1})$. By definition of $u^*_{\pi_m}$, then also:

$$\phi(y_{\pi_1}, \ldots, y_{\pi_m}, x_{\pi_{m+1}}, \ldots, x_{\pi_n}) \leqslant_{\pi_m} \phi(x_{\pi_1}, \ldots, x_{\pi_n}).$$

We may conclude that x is an extortion of order m.

For the only if direction, assume that x is an extortion of G given π. We prove that there is a benign backward induction $f^{(*)}$ in G^* for π^R with $f^{(*)\prime} = x$. In virtue of Fact 4.3, there is a benign backward induction h in G^* given π^R. Now define $f^{(*)}$ in such a way that $f^{(*)}_{\pi_i}(z_{\pi_n}, \ldots, z_{\pi_{i-1}}) = x_{\pi_i}$, if $(z_{\pi_n}, \ldots, z_{\pi_{i-1}}) = (x_{\pi_n}, \ldots, x_{\pi_{i-1}})$, and $f^{(*)}_{\pi_i}(z_{\pi_n}, \ldots, z_{\pi_{i-1}}) = h_{\pi_i}(z_{\pi_n}, \ldots, z_{\pi_{i-1}})$, otherwise. We prove by induction on m, that $f^{(*)}$ is a benign backward induction of order m, for each m with $0 \leqslant m \leqslant n$. The basis is trivial. So assume that $f^{(*)}$ is a backward induction of order $m-1$ in G^* given π^R and consider an arbitrary benign backward induction g of order $m-1$ in G^* given π^R. Let g' be given by $(y_{\pi_n}, \ldots, y_{\pi_1})$. Either $(y_{\pi_n}, \ldots, y_{\pi_{m+1}}) = (x_{\pi_n}, \ldots, x_{\pi_{m+1}})$, or this is not the case. If the latter, it can readily be appreciated that:

$$(g_{\pi_n}, \ldots, g_{\pi_{m+1}}, f^{(*)}_{\pi_m}, \ldots, f^{(*)}_{\pi_1})' = (g_{\pi_n}, \ldots, g_{\pi_{m+1}}, h_{\pi_m}, \ldots, h_{\pi_1})'.$$

Having assumed that h is a benign backward induction, subsequently, $(g_{\pi_n}, \ldots, g_{\pi_1})' \leqslant^*_m (g_{\pi_n}, \ldots, g_{\pi_{m+1}}, h_{\pi_m}, \ldots, h_{\pi_1})'$, and $(g_{\pi_n}, \ldots, g_{\pi_1})' \leqslant^*_m (g_{\pi_n}, \ldots, g_{\pi_{m+1}}, f^{(*)}_{\pi_m}, \ldots, f^{(*)}_{\pi_1})'$. Hence, $f^{(*)}$ is a benign backward induction of order m. In the former case the reasoning is slightly different. Then, $(g_{\pi_n}, \ldots, g_{\pi_1})' = (x_{\pi_n}, \ldots, x_{\pi_{m+1}}, y_{\pi_m}, \ldots, y_{\pi_1})$. It follows that:

$$(g_{\pi_n}, \ldots, g_{\pi_{m+1}}, f^{(*)}_{\pi_m}, \ldots, f^{(*)}_{\pi_1})' = (f^{(*)}_{\pi_n}, \ldots, f^{(*)}_{\pi_1})' = (x_{\pi_n}, \ldots, x_{\pi_1}).$$

In virtue of the induction hypothesis, $(y_{\pi_1}, \ldots, y_{\pi_n})$ is an extortion of order $m-1$ in G given π. As the reasoning takes place under the assumption that x is an extortion in G given π, we also have:

$$\phi(y_{\pi_1}, \ldots, y_{\pi_m}, x_{\pi_{m+1}}, \ldots, x_{\pi_n}) \leqslant_{\pi_m} \phi(x_{\pi_1}, \ldots, x_{\pi_n}).$$

Then, $(x_{\pi_n}, \ldots, x_{\pi_{m+1}}, y_{\pi_m}, \ldots, y_{\pi_1},) \leqslant^*_{\pi_m} (x_{\pi_n}, \ldots, x_{\pi_1})$., by definition of u^*. We may conclude that:

$$(g_{\pi_n}, \ldots, g_{\pi_1})' \leqslant^*_{\pi_m} (g_{\pi_n}, \ldots, g_{\pi_{m+1}}, f^{(*)}_{\pi_m}, \ldots, f^{(*)}_{\pi_1})',$$

signifying that $f^{(*)}$ is a benign backward induction of order m. □

As an immediate consequence of Theorem 4.4 and Fact 4.3 we also have the following result.

COROLLARY 4.5. *Let τ be a finite commitment type. Then, τ-extortions exist for each strategic game and for each ordering of the players.*

4.3 Commitment Order

In the case of unconditional commitments, it is not always favorable to be the first to commit. This is well illustrated by the familiar game *rock-paper-scissors*. If, on the other hand, the players are in a position to make *conditional* commitments in this particular game, moving first is an advantage. Rather, we find that it can never harm to move first in a two-player game with conditional commitments.

THEOREM 4.6. *Let G be a two-player strategic game involving player i. Further let f be an extortion of G in which i commits first, and g an extortion in which i commits second. Then, $g' \leqslant_i f'$.*

PROOF SKETCH. Let f be a conditional extortion in G given π. It suffices to show that there is some conditional extortion h of order 1 in G given π' with $h' = f'$. Assume for a contradiction that there is no such extortion of order 1 in G given π'. Then there must be some $b^* \in A_j$ such that $f' \prec_j (b^*, a)$, for all $a \in A_i$. (Otherwise we could define (g_j, g_i) such that $g_j = f_j(f_i)$, $g_i(g_j) = f_i$, and for any other $b \in A_j$, $g_i(b) = a^*$, where a^* is an action in A_i such that $(b, a^*) \leqslant_j f'$. Then g would be an extortion of order 1 in G given π' with g'.) Now consider a conditional commitment profile h for G and π such that $h_j(a) = b^*$, for all $a \in A_i$. Let further h_i be such that $(a, h_j)' \leqslant_i (h_i, h_j)'$, for all $a \in A_i$. Then, h is an extortion of order 1 in G given π. Observe that $(h_i, h_j)' = (f'_i, b^*)$. Hence, $f' \prec_j h'$, contradicting the assumption that f is an extortion in G given π. □

Theorem 4.6 does not generalize to games with more than two players. Consider the three-player game in Figure 7, with extensive forms as in Figure 8. Here, *Row* and *Mat* have identical preferences. The latter's extortionate powers relative *Col*, however, are very weak if he is to commit first: any conditional commitment he makes puts *Col* in a situation in which she can enforce a payoff of two, leaving *Mat* and *Row* in the cold with a payoff of one. However, if *Mat* is last to commit and *Row* first, then the latter can exploit his strategic powers, threaten *Col* so that she plays left, and guarantee both himself and *Mat* a payoff of two.

4.4 Pareto Efficiency

Another issue concerns the Pareto efficiency of the strategy profiles extortionable through conditional commitments. We say that a strategy profile *s (weakly) Pareto dominates* another strategy profile t if $t \leqslant_i s$ for all players i and $s \not\leqslant_i t$ for some. Moreover, a strategy profile s is *(weakly) Pareto efficient* if it is not (weakly) Pareto dominated by any other strategy profile. We extend this terminology to conditional commitment profiles by saying that a conditional commitment profile f is *(weakly) Pareto efficient* or *(weakly) Pareto dominates* another conditional commitment profile if f' is or does so. We now have the following result.

$$\begin{bmatrix} (0,1,0) & (0,0,0) \\ (0,0,0) & (1,2,1) \end{bmatrix} \qquad \begin{bmatrix} (2,1,2) & (0,0,0) \\ (0,0,0) & (1,2,1) \end{bmatrix}$$

Figure 7: A three-person game.

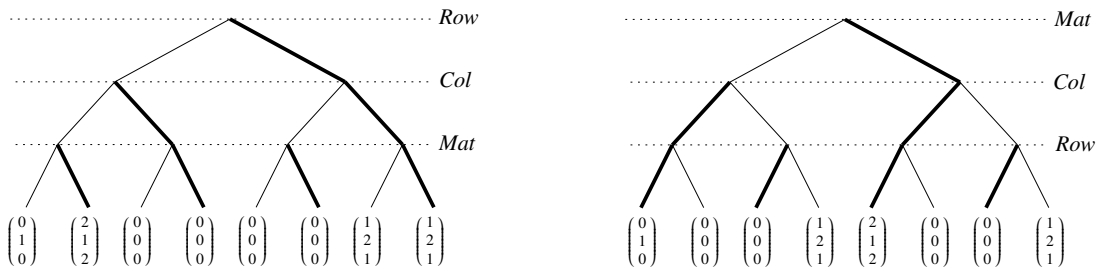

Figure 8: It is not always better to commit early than late, even in the case of conditional or inductive commitments.

THEOREM 4.7. *In each game, Pareto efficient conditional extortions exist. Moreover, any strategy profile that Pareto dominates an extortion is also extortionable through a conditional commitment.*

PROOF SKETCH. Since, in virtue of Fact 4.5, extortions generally exists in each game, it suffices to recognize that the second claim holds. Let s be the strategy profile $(s_{\pi_1}, \ldots, s_{\pi_n})$. Let further the conditional extortion f be Pareto dominated by s. An extortion g with $g' = s$ can then be constructed by adopting all threats of f while promising g'. *I.e.*, for all players π_i we have $g_{\pi_i}(s_{\pi_1}, \ldots, s_{\pi_{i-1}}) = s_i$ and $g_{\pi_i}(t_{\pi_1}, \ldots, t_{\pi_n}) = f_{\pi_i}(t_{\pi_1}, \ldots, t_{\pi_n})$, for all other $t_{\pi_1}, \ldots, t_{\pi_n}$. As s Pareto dominates f', the "threats" of f remain effective as threats of g given that s is being promised. □

This result hints at a difference between (benign) backward induction and extortions. In general, solutions of benign backward inductions can be Pareto dominated by outcomes that are no benign backward induction solutions. Therefore, although every extortion can be seen as a benign backward induction in a larger game, it is not the case that all formal properties of extortions are shared by benign backward inductions in general.

5. OTHER COMMITMENT TYPES

Conditional and unconditional commitments are only two possible commitment types. The definition also provides for types of commitment that allow for committing on commitments, thus achieving a finer adjustment of promises and threats. Similarly, it subsumes commitments on and to mixed strategies. In this section we comment on some of these possibilities.

5.1 Inductive Commitments

Apart from making commitments conditional on the *actions* of the players to commit later, one could also commit on the *commitments* of the following players. Informally, such commitments would have the form of "if you only dare to commit in such and such a way, then I do such and such, otherwise I promise to act so and so."

For a strategic game G and an ordering π of the players, we define the *inductive commitments* of the players inductively. The inductive commitments available to π_1 coincide with the actions that are available to him. An inductive commitment for player π_{i+1} is a function mapping each profile of inductive commitments of players π_1 through π_i to one of his basic actions. Formally we define the type of inductive commitments $(\boldsymbol{F}_{\pi_1}, \ldots, \boldsymbol{F}_{\pi_n}, ')$ such that for each player π_i in a game G and given π:

$$\begin{aligned} \boldsymbol{F}_{\pi_1} &=_{df.} A_{\pi_1}, \\ \boldsymbol{F}_{\pi_{i+1}} &=_{df.} A_{\pi_{i+1}}^{\boldsymbol{F}_{\pi_1} \times \cdots \times \boldsymbol{F}_{\pi_i}}. \end{aligned}$$

Let $\boldsymbol{f}'_{\pi_i} = \boldsymbol{f}_{\pi_i}(\boldsymbol{f}_{\pi_1}, \ldots, \boldsymbol{f}_{\pi_{i-1}})$, for each player π_i and have $\boldsymbol{f}'$ denote the pure strategy profile $(\boldsymbol{f}'_{\pi_1}, \ldots, \boldsymbol{f}'_{\pi_n})$.

Inductive commitments have a greater extortionate power than conditional commitments. To appreciate this, consider once more the game in Figure 5. We found that the strategy profile in which *Row* chooses bottom and *Col* and *Mat* both choose left is not extortionable through conditional commitments. By means of inductive commitments, however, this is possible. Let $\boldsymbol{f}$ be the inductive commitment profile such that $\boldsymbol{f}_{Row}$ is *Row* choosing the bottom row (b), $\boldsymbol{f}_{Col}$ is the column player choosing the left column (l) no matter how *Row* decides, and $\boldsymbol{f}_{Mat}$ is defined such that:

$$\boldsymbol{f}_{Mat}(\boldsymbol{f}_{Row}, \boldsymbol{f}_{Col}) = \begin{cases} r & \text{if } \boldsymbol{f}_{Row} = t \text{ and } \boldsymbol{f}_{Col}(b) = r, \\ l & \text{otherwise.} \end{cases}$$

Instead of showing formally that $\boldsymbol{f}$ is an inductive extortion of the strategy profile (b, l, l), we point out informally how this can be done. We argued that in order to exact a payoff of two by means of a conditional extortion, *Mat* would have to lure *Row* into choosing the bottom row without at the same time putting *Col* in a position to successfully threaten *Row* not to choose top. This, we found, is an impossibility if the players can only make conditional commitments. By contrast, if *Mat* can commit to commitments, he can undermine *Col*'s efforts to threaten *Row* by playing the right matrix, if *Col* were to do so. Yet, *Mat* can still force *Row* to choose the bottom row, in case *Col* desists form making this threat.

As can readily be observed, in any game, the inductive commitments of the first two players to commit coincide with their conditional commitments. Hence, as an immediate consequence of Theorem 4.6, it can never harm to be the first to commit to an inductive commitment in the two player case. Similarly, we find that the game depicted in Figure 7 also serves as an example showing that, in case there are more than two players, it is not always better to commit to an inductive commitment early. In this example the strategic position of *Mat* is so weak if he is to commit first, that even the possibility to commit inductively does not strengthen it, whereas, in a similar fashion as with conditional commitments, *Row* can enforce a payoff of two to both himself and *Mat* if he is the first to commit.

5.2 Mixed Commitments Types

So far we have merely considered commitments to and on pure strategies. A natural extension would be also to consider commitments to and on *mixed strategies*. We distinguish between conditional, unconditional as well as inductive mixed commitments. We find that they are generally quite incomparable with their pure counterparts: in some situations a player can achieve more using a mixed commitment, in another using a pure commitment type. A complicating factor with mixed commitment types is that they

can result in a mixed strategy profile being played. This makes that the distinction between promises and threats, as delineated in Section 4.1, gets blurred for mixed commitment types.

The type of *mixed unconditional commitments* associates with each game G and ordering π of its players the tuple $(\Sigma_{\pi_1}, \ldots, \Sigma_{\pi_n}, id)$. The two-player case has been extensively studied (*e.g.*, [2, 16]). As a matter of fact, von Neumann's famous minimax theorem shows that for two-player zero-sum games, it does not matter which player commits first. If the second player to commit plays a mixed strategy that ensures his security level, the first player to commit can do no better than to do so as well [14].

In the game of Figure 5 we found that, with conditional commitments, *Mat* is unable to enforce an outcome that awards him a payoff of two. Recall that the reason of this failure is that any effort to deter *Row* from choosing the top row is flawed, as it would put *Col* in an excellent position to threaten *Row* not to choose the bottom row. If *Mat* has inductive commitments at his disposal, however, this is a possibility. We now find that in case the players can dispose of unconditional mixed strategies, *Mat* is in a much similar position. He could randomize uniformly between the left and right matrix. Then, *Row*'s expected utility is $2\frac{1}{2}$ if he plays the top row, no matter how *Col* randomizes. The expected payoff of *Col* does not exceed $2\frac{1}{2}$, either, in case *Row* chooses top. By purely committing to the left column, *Col* player entices *Row* to play bottom, as his expected utility then amounts to 3. This ensures an expected utility of three for *Col* as well.

However, a player is not always better off with unconditional mixed commitments than with pure conditional commitments. For an example, consider the game in Figure 2. Using pure conditional commitments, he can ensure a payoff of three, whereas with unconditional mixed commitments $2\frac{1}{2}$ would be the most he could achieve. Neither is it in general advantageous to commit first to a mixed strategy in a three-player game. To appreciate this, consider once more the game in Figure 7. Again committing to a mixed strategy will not achieve much for *Mat* if he is to move first, and as before the other players have no reason to commit to anything other than a pure strategy. This holds for all players if *Row* commits first, *Col* second and *Mat* last, be it that in this case *Mat* obtains the best payoff he can get.

Analogous to conditional and inductive commitments one can also define the types of *mixed conditional* and *mixed inductive* commitments. With the former, a player can condition his mixed strategies on the mixed strategies of the players to commit after him. These tend to be very large objects and, knowing little about them yet, we shelve their formal analysis for future research. Conceptually, it might not be immediately clear how such mixed conditional commitments can be made with credibility. For one, when one's commitments are conditional on a particular mixed strategy being played, how can it be recognized that it was in fact this mixed strategy that was played rather than another one? If this proves to be impossible, how can one know how his conditional commitments is to be effectuated? A possible answer would be, that all depends on the circumstances in which the commitments were made. *E.g.*, if the different agents can submit their mixed conditional commitments to an independent party, the latter can execute the randomizations and determine the unique mixed strategy profile that their commitments induce.

6. SUMMARY AND CONCLUSION

In some situations agents can strengthen their strategic position by committing themselves to a particular course of action. There are various types of commitment, *e.g.*, pure, mixed and conditional. Which type of commitment an agent is in a position in to make essentially depends on the situation under consideration. If the agents commit in a particular order, there is a *tactic* common to making commitments of any type, which we have formalized by means the concept of an *extortion*. This generic concept of extortion can be analyzed *in abstracto*. Moreover, on its basis the various commitment types can be compared formally and systematically.

We have seen that the type of commitment an agent can make has a profound impact on what an agent can achieve in a game-like situation. In some situations a player is much helped if he is in a position to commit conditionally, whereas in others mixed commitments would be more profitable. This raises the question as to the characteristic formal features of the situations in which it is advantageous for a player to be able to make commitments of a particular type.

Another issue which we leave for future research is the computational complexity of finding an extortion for the different commitment types.

7. REFERENCES

[1] A. K. Chopra and M. Singh. Contextualizing commitment protocols. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1345–1352. ACM Press, 2006.

[2] V. Conitzer and T. Sandholm. Computing the optimal strategy to commit to. In *Proceedings of the 7th ACM Conference on Electronic Commerce (ACM-EC)*, pages 82–90. ACM Press, 2006.

[3] J. C. Harsanyi. A simplified bargaining model for the n-person cooperative game. *International Economic Review*, 4(2):194–220, 1963.

[4] R. D. Luce and H. Raiffa. *Games and Decisions: Introduction and Critical Survey*. Wiley, 1957.

[5] J. Nash. Two-person cooperative games. *Econometrica*, 21:128–140, 1953.

[6] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[7] D. Samet. How to commit to cooperation, 2005. Invited talk at the 4th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS).

[8] T. Sandholm and V. Lesser. Leveled-commitment contracting. A backtracking instrument for multiagent systems. *AI Magazine*, 23(3):89–100, 2002.

[9] T. C. Schelling. *The Strategy of Conflict*. Harvard University Press, 1960.

[10] R. Selten. Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit. *Zeitschrift für die gesamte Staatswissenschaft*, 121:301–324, 1965.

[11] M. P. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7(1):97–113, 1999.

[12] M. Tennenholtz. Program equilibrium. *Games and Economic Behavior*, 49:363–373, 2004.

[13] E. van Damme and S. Hurkens. Commitment robust equilibria and endogenous timing. *Games and Economic Behavior*, 15:290–311, 1996.

[14] J. von Neumann and O. Morgenstern. *The Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[15] H. von Stackelberg. *Marktform und Gleichgewicht*. Julius Springer Verlag, 1934.

[16] B. von Stengel and S. Zamir. Leadership with commitment to mixed strategies. CDAM Research Report LSE-CDAM-2004-01, London School of Economics, 2003.