John Grant
Sarit Kraus
Michael Wooldridge
Inon Zuckerman

Manipulating Games by Sharing Information

Abstract. We address the issue of manipulating games through communication. In the specific setting we consider (a variation of Boolean games), we assume there is some set of environment variables, the values of which are not directly accessible to players; the players have their own beliefs about these variables, and make decisions about what actions to perform based on these beliefs. The communication we consider takes the form of (truthful) announcements about the values of some environment variables; the effect of an announcement is the modification of the beliefs of the players who hear the announcement so that they accurately reflect the values of the announced variables. By choosing announcements appropriately, it is possible to perturb the game away from certain outcomes and towards others. We specifically focus on the issue of *stabilisation*: making announcements that transform a game from having no stable states to one that has stable configurations.

Keywords: Boolean games, Nash equilibrium, Communication, Announcement.

1. Introduction

In a Boolean game, [16, 3, 8, 9], each player has a set of Boolean variables under its unique control and is at liberty to assign values to these variables as it chooses. In addition, each player has a goal that it desires to be achieved: the goal is represented as a Boolean formula, which may contain variables under the control of other players. Boolean games have a strategic character because the achievement of one player's goal may depend on the choices of other players. In this work we consider the players to be agents giving the Boolean game an aspect of a multi-agent system. Actually, as we are interested in dealing with communication, we use a special variant of Boolean games: in addition to the variables under the control of the agents, there is an additional set of *environment* variables. An external *principal* knows the values of the environment variables and may announce (truthful) information about them to the agents. The agents then revise their beliefs based on the announcement, and will decide what actions to perform based

Special Issue: Logic and Games
Edited by **Thomas Ågotnes**

Studia Logica (2014) 102: 267–295
DOI: 10.1007/s11225-014-9544-5
© Springer Science+Business Media Dordrecht 2014

on their beliefs. It follows that, by choosing announcements appropriately, the principal can perturb the game away from some possible outcomes and towards others.

We focus on the issue of *stabilisation*: making announcements that transform a game from having no stable states to one that has stable configurations. Stability in this sense is close to the notion of Nash equilibrium in the game-theoretic sense [20]: it means that no agent has any incentive to unilaterally change its choice. However, the difference between our setting and the conventional notion of Nash equilibrium is that an agent's perception of the utility it would obtain from an outcome is dependent on its own beliefs. By changing these beliefs through truthful announcements, we can modify the rational outcomes of the game.

The rationale for focussing on stabilisation is that instability will, in general, be undesirable: apart from anything else, it makes behaviour harder to predict and understand, and introduces the possibility of agents wasting effort by continually modifying their behaviour. It makes sense, therefore, to consider the problem of stabilising multi-agent system behaviour: of modifying an unstable system so that it has equilibrium states, and even further, of modifying the system so that it has socially desirable equilibria. For example, we might consider the principal perturbing a game to ensure an equilibrium that maximises the number of individual agent goals achieved.

Although the model of communication and rational action we consider in the present paper is based on the abstract setting of Boolean games, the issues we investigate using this model — stabilisation, and, more generally, the management of multi-agent systems — are, we believe, of central importance. This is because there is a fundamental difference between a distributed system in which all components are designed and implemented by a single designer, and which can therefore be designed to act in the furtherance of the designer's objectives, and multi-agent systems, in which individual agents will selfishly pursue their own goals. By providing a formal analysis of how communication can be used to perturb the rational actions of agents within a system towards certain outcomes, we provide a foundation upon which future, richer models can be built and investigated.

The plan of the rest of the paper is as follows. In Section 2 we present the basic model that we use for most of the paper. This model allows for complex goals but only simple conjunctive beliefs for the agents. Section 2.2 defines our version of Nash stability. Section 3 introduces simple conjunctive announcements. The major results of the paper in Section 3.2 deal with announcements that stabilise games. Section 3.3 considers measures of optimality for announcements. We then consider extensions to the basic model in Section 4: for example, to allow richer announcements (arbitrary truthful statements about the environment variables), and richer models of belief. Finally, Section 5 considers related work and gives conclusions, including possibilities for further work.

2. The Basic Model

In this section we introduce the basic model of Boolean games that we will work with for most of the paper and define notions of stability for it.

2.1. Components of the Basic Model

Our model is a variation of previous models of Boolean games [16, 3, 8, 9]. The main difference is the addition of a set of *environment variables* whose values are fixed and cannot be changed by the agents. The agents have *beliefs* about the environment variables that may be incorrect, and base their decisions about their choices on their beliefs.

Propositional Logic: Let $\mathbb{B} = \{\top, \bot\}$ be the set of Boolean truth values, with "$\top$" being truth and "$\bot$" being falsity. We will abuse notation a little by using $\top$ and $\bot$ to denote both the syntactic constants for truth and falsity respectively, as well as their semantic counterparts. Let $\Phi = \{p, q, \ldots\}$ be a (finite, fixed, non-empty) vocabulary of Boolean variables, and let $\mathcal{L}$ denote the set of (well-formed) formulae of propositional logic over Φ, constructed using the conventional Boolean operators ("$\wedge$", "$\vee$", "$\rightarrow$", "$\leftrightarrow$", and "$\neg$"), as well as the truth constants "$\top$" and "$\bot$". Where $\varphi \in \mathcal{L}$, we let $vars(\varphi)$ denote the (possibly empty) set of Boolean variables occurring in φ (e.g., $vars(p \wedge q \rightarrow p) = \{p, q\}$).

We will also use a special subset of $\mathcal{L}$. A *simple conjunctive formula* has the form $\ell_1 \wedge \cdots \wedge \ell_k$, where each ℓ_i is a literal, that is, a Boolean variable or its negation. We do not permit both a Boolean variable and its negation to occur in a simple conjunctive formula; hence contradictions are excluded. Such a simple conjunctive formula whose variables are $p_1, \ldots, p_k$ can also be represented as a function $f : \{p_1, \ldots, p_k\} \rightarrow \mathbb{B}$, with $f(\ell_i) = \top$ if $\ell_i = p_i$ and $f(\ell_i) = \bot$ if $\ell_i = \neg p_i$ and we will usually use the latter formulation. A *valuation* is a total function $v : \Phi \rightarrow \mathbb{B}$, assigning truth or falsity to every Boolean variable. We write $v \models \varphi$ to mean that the propositional formula φ is true under, or satisfied by, valuation v, where the satisfaction relation "$\models$" is defined in the standard way. Let $\mathcal{V}$ denote the set of all valuations over Φ. We write $\models \varphi$ to mean that φ is a tautology. We denote the fact that $\models \varphi \leftrightarrow \psi$ by $\varphi \equiv \psi$.

Agents and Variables: The games we consider are populated by a set $Ag = \{1, \ldots, n\}$ of *agents* — the players of the game. Each agent is assumed to have a *goal*, characterised by an $\mathcal{L}$-formula: we write γ_i to denote the goal of agent $i \in Ag$. Agents $i \in Ag$ each *control* a (possibly empty) subset $\Phi_i \subseteq \Phi$. By "control", we mean that i has the unique ability within the game to set the value (either $\top$ or $\bot$) of each variable $p \in \Phi_i$. We will require that $\Phi_i \cap \Phi_j = \emptyset$ for $i \neq j$, but in contrast with other existing models of Boolean games [16, 3], we do *not* require that the sets $\Phi_1, \ldots, \Phi_n$ form a partition of Φ. Thus, we allow for the possibility that some variables are not under the control of any players in the game. Let $\Phi_E = \Phi \setminus (\Phi_1 \cup \cdots \cup \Phi_n)$ be the variables that are not under any agent's control; we call these the *environment* variables. The values of these variables are determined external to the game. We let $v_E : \Phi_E \to \mathbb{B}$ be the function that gives the actual value of the environment variables. When playing a Boolean game, the primary aim of an agent i will be to choose an assignment of values for the variables Φ_i under its control so as to satisfy its goal γ_i. The difficulty is that γ_i may contain variables controlled by other agents $j \neq i$, who will also be trying to choose values for their variables Φ_j so as to get their goals satisfied; and their goals in turn may be dependent on the variables Φ_i. In addition, goal formulae may contain environment variables Φ_E, beyond the control of any agent in the system. A *choice* for agent $i \in Ag$ is a function $v_i : \Phi_i \to \mathbb{B}$, i.e., an allocation of truth or falsity to all the variables under i's control. Let $\mathcal{V}_i$ denote the set of choices for agent i, and let $\mathcal{V}_E$ denote the set of all valuations $v_E : \Phi_E \to \mathbb{B}$ for the set of environment variables Φ_E.

Beliefs: Players in our games are assumed to have possibly incorrect beliefs about the values of the environment variables Φ_E. For the moment, we will assume that the belief a player i has about Φ_E is represented as a simple conjunctive formula over Φ_E, which must include all the variables in Φ_E. For example, suppose that $\Phi_E = \{p, q\}$. Then there are four possible beliefs for a player i:

- $p \wedge q$
  player i believes both that p and q are true;
- $p \wedge \neg q$
  player i believes that p is true and q is false;
- $\neg p \wedge q$
  player i believes that p is false and q is true; and

- $\neg p \wedge \neg q$
  player i believes that both p and q are false.

Thus, in our simple model of belief, formulae such as $p \vee q$ and $p \rightarrow q$ are not allowed.

We will model the beliefs of an agent $i \in Ag$ via the corresponding function $\beta_i : \Phi_E \rightarrow \mathbb{B}$. Thus, $\beta_i(p) = b$ indicates agent i's belief that $p \in \Phi_E$ has the value b (where $b \in \mathbb{B}$).

This is, of course, a very simple model of belief, and many alternative richer models of belief could be used instead. Later in this paper, we will consider some such richer models of belief.

Outcomes: An *outcome* is a collection of choices, one for each agent. Formally, an outcome is a tuple $(v_1, \ldots, v_n) \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$. When taken together with the valuation v_E for the environment variables, an outcome uniquely defines an overall valuation for all the variables in Φ. We write $(v_1, \ldots, v_n, v_E) \models \varphi$ to mean that the valuation defined by the outcome $(v_1, \ldots, v_n)$ taken together with v_E satisfies formula $\varphi \in \mathcal{L}$. A belief function β_i together with an outcome $(v_1, \ldots, v_n)$ also defines a unique valuation for Φ, and we will write $(v_1, \ldots, v_n, \beta_i)$ to mean the valuation obtained from the choices $v_1, \ldots, v_n$ together with the values for the variables in Φ_E defined by β_i. Observe that we could have $(v_1, \ldots, v_n, \beta_i) \models \gamma_i$ (agent i believes that its goal γ_i is achieved by outcome $(v_1, \ldots, v_n)$) while $(v_1, \ldots, v_n, v_E) \not\models \gamma_i$ (in fact, it is not). Let $succ(v_1, \ldots, v_n, v_E)$ denote the set of agents whose goals are actually achieved by outcome $(v_1, \ldots, v_n)$, that is:

$$succ(v_1, \ldots, v_n, v_E) = \{i \in Ag \mid (v_1, \ldots, v_n, v_E) \models \gamma_i\}.$$

Costs: Intuitively, the actions available to agents correspond to setting variables to true or false. We assume that these actions have *costs*, defined by a *cost function* $c : \Phi \times \mathbb{B} \rightarrow \mathbb{R}_{\geq}$, so that $c(p, b)$ is the marginal cost of assigning variable $p \in \Phi$ the value $b \in \mathbb{B}$ (where $\mathbb{R}_{\geq} = \{x \in \mathbb{R} \mid x \geq 0\}$). Note that if an agent has multiple ways of getting its goal achieved, then it will prefer to choose one that minimises costs; and if an agent cannot get its goal achieved, then it simply chooses to minimise costs. However, cost reduction is a *secondary* consideration: an agent's primary concern is goal achievement.

To keep the model simple, we will assume that the cost of setting a variable to $\bot$ is 0; this makes sense if we think of Boolean variables as actions, and cost as the corresponding marginal cost of performing a particular action. In this case setting a variable to be $\top$ means performing an action, and hence incurring the corresponding marginal cost, while setting it to $\bot$ corresponds to doing nothing, and hence incurring no cost.

Boolean Games: A Boolean game, G, is a $(3n+4)$-tuple:

$$G = \langle Ag, \Phi, \underbrace{\Phi_1, \ldots, \Phi_n}_{\substack{\text{controlled}\\\text{variables}}} \underbrace{\gamma_1, \ldots, \gamma_n}_{\text{goals}}, \underbrace{\beta_1, \ldots, \beta_n}_{\text{beliefs}}, c, v_E \rangle,$$

where:

- $Ag = \{1, \ldots, n\}$ is a set of agents — the *players* of the game;
- $\Phi = \{p, q, \ldots\}$ is the (finite) set of Boolean variables;
- $\Phi_i \subseteq \Phi$ is the set of Boolean variables under the unique control of player $i \in Ag$;
- $\gamma_i \in \mathcal{L}$ is the goal of agent $i \in Ag$;
- β_i represents the beliefs of agent $i \in Ag$;
- $c : \Phi \times \mathbb{B} \to \mathbb{R}_{\geq}$ is the cost function; and
- $v_E : \Phi_E \to \mathbb{B}$ is the (fixed) valuation function for the environment variables.

For now, as we are dealing with Boolean games with simple conjunctive beliefs, it is convenient to represent β_i as a belief function $\beta_i : \Phi_E \to \mathbb{B}$. However, it should be understood that this function is the representation of a simple conjunctive formula over Φ_E; since these two representations are directly equivalent for simple conjunctive formulae, in what follows we will use whichever representation is more convenient for the task at hand.

2.2. Stability

Now we are ready to define the notion of equilibrium that we use throughout the paper. We call it *Nash stability* as it is a variation of a concept with the same name that was defined in [13]. Nash stability is, in turn, derived from the well-known notion of *pure strategy Nash equilibrium* from non-cooperative game theory [20].

Subjective Utility: We now introduce a model of utility for our games. While we find it convenient to define numeric utilities, it should be clearly understood that utility is not assumed to be transferable: it is simply a numeric way of capturing an agent's preferences. The basic idea is that an agent will strictly prefer all outcomes in which it gets its goal achieved over all outcomes where it does not; and secondarily, will prefer to minimise costs. Utility functions as we define them directly capture such preferences.

When an agent makes a choice, it intuitively makes a calculation of the utility that it might obtain from that choice. However, this calculation is *subjective*, in the sense that the agent's beliefs may be wrong, and hence its judgement about the utility it will obtain from making a choice may be wrong. We let $u_i(v_1, \ldots, v_n)$ denote the utility that agent i believes it would obtain if agent j $(1 \leq j \leq n)$ made choice v_j. We define it formally as follows. First, we let $c_i(v_i)$ denote the marginal cost to agent i of choice $v_i \in \mathcal{V}_i$:

$$c_i(v_i) = \sum_{p \in \Phi_i} c(p, v_i(p))$$

The highest possible cost for agent i, which we write as μ_i, occurs when agent i sets all its variables to $\top$. We then define the subjective utility that i would obtain from choices $v_1, \ldots, v_n$ as the negative of the cost to the agent if its goal is not satisfied; otherwise it is the positive difference (+1) between the highest possible cost and the actual cost.

$$u_i(v_1, \ldots, v_n) = \begin{cases} 1 + \mu_i - c_i(v_i) & \text{if } (v_1, \ldots, v_n, \beta_i) \models \gamma_i \\ -c_i(v_i) & \text{otherwise.} \end{cases}$$

Sometimes we will want to be explicit about the beliefs an agent is using when computing subjective utility, in which case we will write $u_i(v_1, \ldots, v_n, \beta_i)$ to mean the utility agent i will get assuming the belief function β_i.

Thus a player receives positive utility if its goal is satisfied, and negative utility if its goal is not satisfied. In order to maximize utility, each agent will try to satisfy its goal by adopting a valuation that does so with minimal cost; if it cannot satisfy its goal, it adopts the valuation of minimal cost, namely setting all its variables to $\bot$. It is important to note that in this definition the value of an agent's utility is critically dependent on its beliefs β_i.

Nash Stability: The basic idea of Nash stability, as with (pure strategy) Nash equilibrium [20], is that an outcome is said to be Nash stable if no agent within it would prefer to make a different choice, assuming every other agent stays with its choice. However, the difference between Nash stability and the conventional notion of Nash equilibrium is that an agent i in our setting will compute its utility — and hence make its choice — based on its beliefs β_i. We say an outcome $(v_1, \ldots, v_i, \ldots, v_n)$ is *individually stable for agent i* if $\nexists v_i' \in \mathcal{V}_i$ such that $u_i(v_1, \ldots, v_i', \ldots, v_n) > u_i(v_1, \ldots, v_i, \ldots, v_n)$. We then say that an outcome $(v_1, \ldots, v_n)$ is *Nash stable* if $(v_1, \ldots, v_n)$ is individually stable for all players $i \in Ag$. We denote the Nash stable outcomes of a game G by $NE(G)$. As with pure strategy Nash equilibria, it may be that $NE(G) = \emptyset$; in this case we call G *unstable*. If a game is unstable, then for every possible outcome $(v_1, \ldots, v_n)$ of the game, some player would do better to make an alternative choice. We say that such a player has a *beneficial deviation*.

Dependencies: Recall now our earlier statement that the achievement of an agent's goal may depend on the actions of other agents. We will now make this idea formal, in the notion of a *dependency graph* [4]. A dependency graph is a digraph in which the vertex set is the set of players of a game, and where an edge from one player to another indicates that the utility that the source player gets may depend on the choices of the destination player. Formally, a dependency graph for a Boolean game G is a digraph $D_G = (V, E)$, with vertex set $V = Ag$ and edge set $E \subseteq Ag \times Ag$ defined as follows:

$(i, j) \in E$
iff
$\exists(v_1, \ldots, v_j, \ldots, v_n) \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_j \times \cdots \times \mathcal{V}_n$ and $v'_j \in \mathcal{V}_j$ such that
$u_i(v_1, \ldots, v_j, \ldots, v_n) \neq u_i(v_1, \ldots, v'_j, \ldots, v_n)$.

In words, $(i, j) \in E$ if there is some circumstance under which a choice made by agent j can affect the utility obtained by agent i. Where $D_G = (V, E)$, we will subsequently abuse notation and write $(i, j) \in D_G$ to mean $(i, j) \in E$.

Proposition 6 of [4] gives a sufficient condition for the existence of a Nash stable outcome: namely, if the irreflexive portion of D_G is acyclic then $NE(G) \neq \emptyset$. As was shown in [4] this condition is not necessary; for instance, if two agents have the same goal, a cycle between them is irrelevant. As we now show, in general, the problem of checking for acyclicity is computationally complex.

PROPOSITION 1. *Given a game G and agents i, j in G, the problem of determining whether $(i, j) \in D_G$ is* NP*-complete.*

PROOF. Membership is by "guess-and-check". For hardness, we reduce SAT. Let φ be a SAT instance. Create two agents, 1 and 2, let $\gamma_1 = \varphi \wedge z$, where z is a new variable, and let $\gamma_2 = \top$. Let $\Phi_1 = vars(\varphi)$ and $\Phi_2 = \{z\}$. All costs are 0. We now ask whether 1 is dependent on 2; we claim the answer is "yes" iff φ is satisfiable:

($\rightarrow$) Observe that the only way player 1 could obtain different utilities from two outcomes varying only in the value of z (the variable under the control of 2) is if $\varphi \wedge z$ were true in one outcome and false in the other. The outcome satisfying $\varphi \wedge z$ is then witness to the satisfiability of φ.

($\leftarrow$) If agent 1 gets the same utility for all choices as well as for either choice for z then $\varphi \wedge z$ is not satisfiable, hence φ is not satisfiable. ■

Next we show that if not just the beliefs of the agents but also their goals are in simple conjunctive form, the computational cost is reduced.

PROPOSITION 2. *In a game where all players have simple conjunctive beliefs and goals, we have* $(i, j) \in D_G$ *iff* $vars(\gamma_i) \cap \Phi_j \neq \emptyset$. *It follows that, for a game* G *with simple conjunctive beliefs and goals, computing* D_G *can be done in polynomial time: we simply have to check for each pair of agents* $\{i, j\} \subseteq Ag$ $(i \neq j)$ *whether or not* $vars(\gamma_i) \cap \Phi_j \neq \emptyset$, *which is trivially computed in polynomial time.*

To illustrate the ideas we have introduced above, we now present a small (and slightly playful) example.

EXAMPLE 1. Consider the following scenario:

> Bob likes Alice, and he believes Alice likes him. Although Bob doesn't like going to the pub usually, he would want to be there if Alice likes him and Alice was there also. Alice likes going to the pub, but in fact she doesn't like Bob: she wants to go to the pub only if Bob isn't there.

We formalise this example in our setting as follows. The atomic propositions are:

ALB – Alice likes Bob;

PA – Alice goes to the pub, and

PB – Bob goes to the pub.

We have:

$\Phi_A = \{PA\}$ (Alice can determine whether she goes to the pub);

$\Phi_B = \{PB\}$ (Bob can determine whether he goes to the pub); and

$\Phi_E = \{ALB\}$ (the environment determines whether Alice likes Bob).

We also have $v_E(ALB) = \bot$ (in fact, Alice does not like Bob), and $\beta_A(ALB) = \bot$ (Alice believes she does not like Bob), but $\beta_B(ALB) = \top$ (Bob believes Alice likes him — poor deluded Bob!).

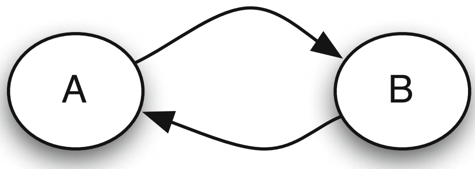

Figure 1. The dependency graph for Example 1.

For both agents $i \in \{A, B\}$ we have $c(Pi, \top) = 10$ (take this to be the cost of a couple of drinks in the pub), while $c(Pi, \bot) = 0$ (staying at home costs nothing).

Alice's goal is simply to avoid Bob:

$$\gamma_A = \neg(PA \leftrightarrow PB).$$

However, Bob's goal is that Alice likes him, and is in the pub with him:

$$\gamma_B = ALB \wedge (PB \leftrightarrow PA).$$

Now, it is easy to see that the game has no Nash stable state:

- If $PA = PB = \bot$, then Alice would benefit by setting $PA = \top$, thereby achieving her goal.
- If $PA = \bot$ and $PB = \top$, then Alice gets her goal achieved but Bob does not; he would do better to set $PB = \bot$.
- If $PA = \top$ and $PB = \bot$, then, again Alice gets her goal achieved but Bob does not; he would do better to set $PB = \top$.
- Finally, if $PA = \top$ and $PB = \top$, then Bob gets his goal achieved but Alice does not; she would do better to set $PA = \bot$.

The irreflexive portion of the dependency graph for this example is shown in Figure 1: observe that both players are dependent upon each other, and so we have a cycle in the dependency graph.

3. Simple Conjunctive Announcements

Let us now return to the motivation from the introduction of the paper: namely, that a principal makes announcements about the values of the environment variables in order to modify the behaviour of agents within the system. We now begin our investigation of this issue by assuming that announcements take the form of simple conjunctions; we will call them *simple* *conjunctive announcements.* We start by distinguishing between two types of announcements. This is followed by a study of the issue of stabilizing a Boolean game by such an announcement. Finally, we consider the problem of finding an optimal stabilizing announcement.

3.1. Types of Announcements

We will consider two types of announcements: uniform and nonuniform. By uniform we mean that the announcements are the same for all the agents. Since an announcement (in this section) has simple conjunctive form, its effect is to reveal the truth values of some, possibly all, environment variables. As the agent beliefs for now are also in simple conjunctive form, the effect of an announcement is for the agents to replace false beliefs by true beliefs. We emphasise that announcements must be *truthful*: the principal cannot lie about the values of the variables.

Uniform Announcements: Formally, we model a uniform announcement as a subset $\alpha \subseteq \Phi_E$ ($\alpha \neq \emptyset$), with the intended meaning that, if the principal makes this announcement, then the actual value of every variable $p \in \alpha$ becomes common knowledge within the game. The effect of such an announcement α on an agent's belief function $\beta_i : \Phi_E \to \mathbb{B}$ is to transform it to a new belief function $\beta_i \oplus \alpha$, defined as follows:

$$\beta_i \oplus \alpha(p) = \begin{cases} v_E(p) & \text{if } p \in \alpha \\ \beta_i(p) & \text{otherwise.} \end{cases}$$

With a slight abuse of notation, where G is a game and α is a possible announcement in G, we will write $G \oplus \alpha$ to denote the game obtained from G by replacing every belief function β_i in G with the belief function $\beta_i \oplus \alpha$. Observe that, given a game G and announcement α from G, computing $G \oplus \alpha$ can be done in polynomial time.

Notice that we can view a uniform announcement α either set theoretically (as a subset of Φ_E, the idea being that the value of every member of α is revealed), or else as a conjunctive formula:

$$\left(\bigwedge_{\{p \in \alpha | v_E(p) = \top\}} p \right) \wedge \left(\bigwedge_{\{q \in \alpha | v_E(p) = \bot\}} \neg q \right)$$

We will switch between these two views as we find it convenient.

Nonuniform Announcements: We model nonuniform announcements as functions $\alpha : Ag \to 2^{\Phi_E}$, with the intended interpretation that after making an announcement α, an agent i comes to know the value of the environment variables $\alpha(i)$. Thus, with a non-uniform announcement, the principal can reveal different pieces of information to different agents. As with uniform announcements, the effect of a nonuniform announcement α on an agent is to transform its belief function β_i to a new function $\beta_i \oplus \alpha$, which in this case is defined as follows:

$$\beta_i \oplus \alpha(p) = \begin{cases} v_E(p) & \text{if } p \in \alpha(i) \\ \beta_i(p) & \text{otherwise.} \end{cases}$$

The *size* of a nonuniform announcement α, is denoted (with a small abuse of notation) by $|\alpha|$ and is defined as: $|\alpha| = \sum_{i \in Ag} |\alpha(i)|$.

Recall that a key concern of our work is announcements that stabilise games. So, we return to Example 1 where the game had no Nash stable state.

Example 2. Suppose that Alice's friend, the principal, announces $\{ALB\}$ to Bob; that is, she tells Bob that Alice does not in fact like him. Notice that announcing $\{ALB\}$ to Bob means revealing the value of the environment variable ALB, which can also be viewed as announcing to Bob the formula $\neg ALB$. After the announcement Bob updates his belief accordingly. At this point, Bob no longer has any possibility to achieve his goal $ALB \wedge (PB \leftrightarrow PA)$, and his optimal choice is to minimise costs by not going to the pub. Given that Bob stays at home, Alices's optimal choice is to go to the pub. The outcome where $PA = \top$, $PB = \bot$ (Alice goes to the pub but Bob stays at home) is Nash stable. Thus, announcing $\alpha = \{ALB\}$ to Bob only is a nonuniform stabilizing announcement. As Alice already believes α, the same effect could have been achieved by the uniform announcement in which $\{ALB\}$ was announced to both Bob and Alice.

3.2. Simple Conjunctive Announcements that Stabilize Games

The last example showed a situation where an announcement (either uniform or nonuniform) changed a Boolean game from one without a Nash stable state to one that is Nash stable. In this subsection we will start by considering uniform announcements. We will say that an announcement α is *stabilising* if $NE(G \oplus \alpha) \neq \emptyset$ (we do not require that $NE(G) = \emptyset$). Let $\mathcal{S}(G)$ be the set of uniform simple conjunctive stabilising announcements for G:

$$\mathcal{S}(G) = \{\alpha \subseteq \Phi_E \mid NE(G \oplus \alpha) \neq \emptyset\}.$$

From the point of view of the principal, the obvious decision problem relating to stabilisation is as follows:

Given a game G, does there exist some announcement α over G that stabilises G, i.e., is it the case that $\mathcal{S}(G) \neq \emptyset$?

We have the following:

PROPOSITION 3. *The problem of checking whether a game G can be stabilised by a uniform simple conjunctive announcement, (i.e., whether $\mathcal{S}(G) \neq \emptyset$), is Σ_2^p-complete; this holds even if all costs are* 0.

PROOF. Membership is by the following algorithm: Guess an $\alpha \subseteq \Phi_E$ and an outcome $(v_1, \ldots, v_n)$, and verify that $(v_1, \ldots, v_n)$ is a Nash stable outcome of $G \oplus \alpha$. Guessing can clearly be done in non-deterministic polynomial time, and verification is a co-NP computation. For hardness, we reduce the problem of checking whether a Boolean game as defined in [3] has a Nash equilibrium; this problem was shown to be Σ_2^p-complete in [3]. Given a conventional Boolean game, we map the agents, goals, and controlled variables to our setting directly; we then create one new Boolean variable, call it z, and set $\Phi_E = \{z\}$. Let $v_E(z) = \top$ and $\beta_i(z) = \top$ for all agents i. Now, we claim that the system can be stabilised iff the original game has a Nash equilibrium; the only announcement that can be made is $\alpha = \{z\}$, which does not change the system in any way; the Nash stable states of the game $G \oplus \alpha$ will thus be exactly the Nash equilibria of the original game. ■

Another obvious question is what properties announcements have. While this is not the primary subject of the present paper, it is nevertheless worth considering. We have the following:

PROPOSITION 4. *Stability is not monotonic through announcements. That is, there exist games G and announcements α_1, α_2 over G such that $G \oplus \alpha_1$ is stable but $(G \oplus \alpha_1) \oplus \alpha_2$ is not.*

PROOF. Consider the following example (a variant of the Alice and Bob example introduced earlier). Let G be the game with:

- $Ag = \{1, 2\}$;
- $\Phi = \{p, q, r, s\}$;
- $\Phi_1 = \{p\}$;
- $\Phi_2 = \{q\}$;
- $\Phi_E = \{r, s\}$;
- $\beta_1(r) = \top$;
- $\beta_1(s) = \bot$;

- $\beta_2(r) = \bot$;
- $\beta_2(s) = \top$;
- $v_E(r) = \bot$;
- $v_E(s) = \top$ (so agent 2's belief is correct);
- $\gamma_1 = (r \vee s) \wedge (p \leftrightarrow q)$;
- $\gamma_2 = \neg(p \leftrightarrow q)$; and
- $c(p, \top) = c(q, \top) = 1$.

Now, G is unstable, by a similar argument to Example 1. Announcing $\neg r$ will stabilise the system, again by a similar argument to Example 1. However, it is easy to see that $(G \oplus \{\neg r\}) \oplus \{s\}$ is unstable: intuitively, in $(G \oplus \{\neg r\})$, agent 1 does not believe that it can get its goal achieved, because it believes both r and s are false, so it prefers to minimise costs by setting $p = \bot$, leaving agent 2 free to get its goal achieved by setting $q = \top$. However, in $(G \oplus \{r\}) \oplus \{s\}$, because agent 1 now believes again, as in the case of G, that there is some possibility to get its goal achieved, the system is unstable. ■

Let us say an announcement $\alpha \subseteq \Phi$ is *relevant* for an agent i if the announcement refers to variables that occur in the goal of i, that is, if $\alpha \cap \mathit{vars}(\gamma_i) \neq \emptyset$. Call α *irrelevant* if it is not relevant for any agent. Clearly, if α is irrelevant w.r.t. G then $\mathit{NE}(G) = \mathit{NE}(G \oplus \alpha)$.

Recall now our earlier statement about the dependency graph D_G for a game and the condition that if the irreflexive portion of D_G is acyclic, then $\mathit{NE}(G) \neq \emptyset$. Consider a game G where $\mathit{NE}(G) = \emptyset$. By this condition the irreflexive portion of D_G must have a cycle. Now, suppose the principal can make an announcement that *breaks* all such cycles; such an announcement α will stabilise the system and must be such that the irreflexive portion of $D_{G\oplus\alpha}$ is acyclic. This suggests an approach to stabilizing systems through announcements: try to find an announcement α such that the irreflexive portion of $D_{G\oplus\alpha}$ is acyclic.

We illustrate this situation with the following example.

Example 3. Consider the following game G:

- $Ag = \{1, 2, 3\}$;
- $\Phi = \{p_1, p_2, p_3, q_1, q_2, q_3\}$;
- $\Phi_i = \{p_i\}$ (hence the q's are the environment variables);
- $\gamma_1 = p_1 \vee p_2 \vee q_1$;

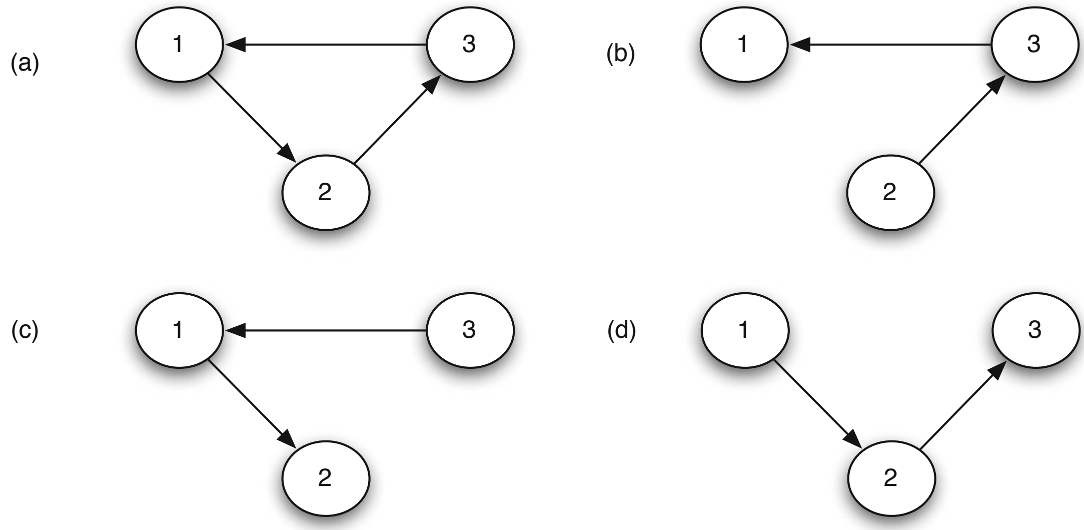

Figure 2. The dependency graph for Example 3. Part (a) shows the original dependency graph; part (b) shows the graph after announcement α_1, while parts (c) and (d) show the dependency graph after announcements α_2 and α_3 respectively.

- $\gamma_2 = p_2 \vee p_3 \vee q_2$;
- $\gamma_3 = p_3 \vee p_1 \vee q_3$;
- $c(p_i, \top) = 1$ for all $i \in \{1, 2, 3\}$;
- $v_E(q_i) = \top$ for all $i \in \{1, 2, 3\}$; and finally,
- $\beta_i(q_j) = \bot$ for all $i, j \in \{1, 2, 3\}$.

The system is unstable: for example, the outcome in which all variables take the value $\bot$ is unstable because agent 1 could benefit by setting $p_1 = \top$. The irreflexive portion of the dependency graph for this example is illustrated in Figure 2(a).

Observe, however, that any of the following announcements would serve to stabilise the system:

- $\alpha_1 = \{q_1\}$;
- $\alpha_2 = \{q_2\}$; or
- $\alpha_3 = \{q_3\}$.

For example, if announcement α_1 is made, then agent 1 will believe its goal will be achieved, and so only needs to minimise costs — it need not be concerned with what agent 2 does with p_2, so it sets $p_1 = \bot$. In this case, agent 3's best response is setting $p_3 = \top$ (thereby achieving its goal), and agent 2 can set $p_2 = \bot$, minimising its cost. This outcome is stable. Identical arguments show that α_2 or α_3 would also stabilise the system. Note how in Figure 2(a) the irreflexive portion of the original D_G contains the edges $(1,2),(2,3),(3,1)$ creating a cycle. The announcement α_1 breaks the dependency and hence removes the edge $(1,2)$ making the irreflexive portion of the new digraph acyclic. Similarly, α_2 removes edge $(2,3)$ and α_3 removes edge $(3,1)$. Hence, if at least one of the environment variables is true, the principal can stabilise the system.

Recall from Proposition 2 that for games with goals in simple conjunctive form, we can easily identify the dependencies between agents. The next question is how to break these dependencies. As in Example 3, the basic idea is to modify an agent's beliefs so that it no longer believes its optimal choice is dependent on the choices of others. We do this by convincing the agent that its goal is either guaranteed to be achieved (in which case its optimal choice is to minimise costs), or else cannot be achieved (in which case, again, the optimal choice is again simply to minimise costs). The difficulty with this approach is that we need to be careful, when making such an announcement, not to change the beliefs of other agents so that the dependency graph contains a new cycle; nonuniform announcements will enable us to manipulate the beliefs of individual agents without affecting those of others.

Where γ_i is a goal for some agent in a game G and α is an announcement, let $\tau(\gamma_i, \alpha)$ denote the formula obtained from γ_i by systematically replacing each variable $p \in \Phi_E$ by $\beta_i \oplus \alpha(p)$. We will say that α *settles* a goal γ_i if $\tau(\gamma_i, \alpha) \equiv \top$ or $\tau(\gamma_i, \alpha) \equiv \bot$. Intuitively, α settles γ_i if the result of making the announcement α is that i believes its goal is guaranteed to be true or is guaranteed to be false. So if G is a game with cyclic dependency graph $D_G = (V, E)$, containing an edge (i, j) such that the irreflexive portion of $E \setminus \{(i, j)\}$ is acyclic, and such that γ_i can be settled by some (nonuniform) announcement α, then G can be stabilised. For games with simple conjunctive goals we can check this condition in polynomial time. For games in general, of course, checking the conditions will be harder.

3.3. Measures of Optimality for Announcements

Apart from asking whether some stabilising announcement exists, it seems obvious to consider the problem of finding an "optimal" stabilising announcement. There are many possible notions of optimality that we might consider, but here, we define just three.

Minimal Stabilising Announcements: The most obvious notion of optimality we might consider for announcements is that of *minimising size.* That is, we want an announcement α^* satisfying:

$$\alpha^* \in \arg\min_{\alpha \in \mathcal{S}(G)} |\alpha|.$$

PROPOSITION 5. *The problem of computing the size of the smallest stabilising uniform (respectively nonuniform) simple conjunctive announcement is in* $\mathrm{FP}^{\Sigma_2^p[\log_2|\Phi|]}$ *(resp.* $\mathrm{FP}^{\Sigma_2^p[\log_2|\Phi\times Ag|]}$*).*

PROOF. We give the proof for uniform announcements; the case for nonuniform announcements is similar. Observe that the following problem, which we refer to as P, is Σ_2^p-complete using similar arguments to Proposition 3: *Given a game G, announcement α for G and $n \in \mathbb{N}$ ($n \leq |\Phi_E|$), does there exist a uniform stabilising announcement α' for G, where $\alpha \subseteq \alpha'$, such that $|\alpha'| \leq n$?* It then follows that, for uniform announcements, determining the size of the smallest stabilising announcement can be computed with $\log_2|\Phi|$ queries to an oracle for P using binary search (cf. [21, pp.415–418]). ■

PROPOSITION 6. *The problem of computing a smallest stabilising uniform (respectively nonuniform) simple conjunctive announcement is in* $\mathrm{FP}^{\Sigma_2^p[|\Phi|]}$ *(respectively* $\mathrm{FP}^{\Sigma_2^p[|Ag\times\Phi|]}$*).*

PROOF. Compute the size s of the smallest announcement using the procedure of Proposition 5. Then we build a stabilising announcement α^* by dynamic programming: A variable S will hold the "current" announcement, with $S = \emptyset$ initially. Iteratively consider each variable $p \in \Phi_E$ in turn, invoking the oracle for P to ask whether there exists a stabilising announcement for G of size s using the partial announcement $S \cup \{p\}$; if the answer is yes, then we set $S = S \cup \{p\}$. We then move on to the next variable in Φ_E. We terminate when $|S| = s$. In this case, S will be a stabilising announcement of size s, i.e., it will be a smallest stabilising announcement. The overall number of queries to the Σ_2^p oracle for P is $|\Phi| + \log_2|\Phi|$, i.e., $O(|\Phi|)$. ■

Goal Maximising Announcements: We do not have transferable utility in our setting, so it makes no sense to directly introduce a measure of social welfare (normally defined for an outcome as the sum of the utilities of the players in that outcome). However, a reasonable proxy for social welfare in our setting is to count the number of goals that are achieved in the "worst"

Nash stable outcome. Formally, we want an announcement α^* satisfying:

$$\alpha^* \in \arg\max_{\alpha \in \mathcal{S}(G)} \min \quad \{succ(v_1, \ldots, v_n, v_E) \mid (v_1, \ldots, v_n, v_E) \in NE(G \oplus \alpha)\}.$$

Objective Satisfying Announcements: A third possibility, considered in [9], is the idea of modifying a game so that a particular objective is achieved in equilibrium, where the objective is represented as a formula $\Upsilon \in \mathcal{L}$. Formally, given a game G and an objective $\Upsilon \in \mathcal{L}$, we seek an announcement $\alpha^* \in \mathcal{S}(G)$ such that:

$$\forall (v_1, \ldots, v_n) \in NE(G \oplus \alpha^*) : (v_1, \ldots, v_n, v_E) \models \Upsilon.$$

Unique Equilibria: As described above, the main aim of this paper is to consider the issue of stabilising Boolean games through announcements. However, we may have games with more than one Nash equilibrium, which then presents the players with a *coordination* problem: which equilibrium should they choose? The following example shows that the same technique described in this paper can be used to ensure that a game has a *unique* equilibrium.

Example 4. Suppose we have a game with:

- $Ag = \{1, 2\}$;
- $\Phi_1 = \{p\}$;
- $\Phi_2 = \{q\}$;
- $\Phi_E = \{r\}$;
- $\gamma_1 = p \vee r$;
- $\gamma_2 = p \wedge q$;
- $c(p, b) = 0$ for all $p \in \Phi,\ b \in \mathbb{B}$;
- $v_E(r) = \bot$;
- $\beta_1(r) = \top$; and
- $\beta_2(r) = \top$.

Essentially, player 1 (incorrectly) believes its goal is satisfied by virtue of the fact that the environment variable r is true, and so is indifferent between setting p to be $\top$ or $\bot$. There are thus three equilibria in the game, corresponding to the formulae $p \wedge q$, $\neg p \wedge q$, and $\neg p \wedge \neg q$. (The formula $p \wedge \neg q$ is not an equilibrium because player 2 could improve the outcome for himself by setting $q = \top$). However, if the principal announces the true value of r, then player 1 will believe the only way its goal will be achieved is by setting $p = \top$; player 2 will then be able to achieve its goal by setting $q = \top$. Thus the game will then have a unique Nash equilibrium that satisfies both players' goals.

4. Extensions and Refinements

In this section, we present and discuss a number of possible refinements and extensions to the basic model of games, announcements, and updates that we presented above. All of these extensions and refinements have a common theme: namely, that they are concerned with going beyond the very simple model of beliefs, going beyond the simple associated model of announcements, and going beyond the simple model of belief update that we presented earlier.

4.1. Complex Goals, Beliefs, and Announcements

Up to this point we have dealt with Boolean games where beliefs and announcements are simple conjunctive formulae. In this section we remove that restriction and allow arbitrary propositional formulae for both beliefs and announcements. Recall first that while the goals are formulae that may contain any of the variables of Φ, the beliefs and announcements are restricted to the environment variables Φ_E. We will represent both an agent's belief and an announcement in what we call minimal disjunctive form, as we define it below. In analogy with the concept of simple conjunctive form, we say that a formula is in *simple disjunctive form* (also known as a clause in other contexts) if it is a disjunction of literals. Where we write "disjunction", this should be understood as an abbreviation of "formula in minimal disjunctive form". A set of disjunctions S is in *minimal disjunctive form* if the following two conditions hold:

1. If $d \in S$ then $S \setminus d \not\models d$, and
2. If $d \in S$ then there is no proper subformula d' of d such that $S \models d'$.

If a set of disjunctions S is not in minimal disjunctive form, it can always be changed to an equivalent set in minimal disjunctive form by omitting superfluous (for equivalence) formulae and replacing every formula by a subformula, if there is one, that is implied by S. These steps may have to be done multiple times. We write $md(S)$ for the minimal disjunctive form of S. Clearly, $md(S)$ is uniquely defined for every S.

EXAMPLE 5. Let $\Phi_E = \{p, q, r\}$. Then:

- $\delta_1 = \{p \vee q, \neg p \vee r\}$ is in minimal disjunctive form.
- $\delta_2 = \{p \vee q, \neg p \vee r, q \vee r\}$ is not in minimal disjunctive form because it violates the first condition ($q \vee r$ is superfluous).
- $\delta_3 = \{p \vee q \vee r, p \vee q \vee \neg r\}$ is not in minimal disjunctive form because the two disjunctions can be replaced by $p \vee q$.
- $\delta_4 = \{p \vee q, \neg p \vee r, p \vee q \vee r\}$ violates both conditions for minimal disjunctive form.

The revision of an agent's belief given an announcement was easy when only simple conjunctive formulae were allowed. We just replaced in the agent's belief set the incorrect truth value for any environment variable disclosed in the announcement. We wrote β_i for the belief of agent i, α for the announcement and $\beta_i \oplus \alpha$ for the updated belief of the agent. As we will be dealing with only one agent for now, we will drop the subscript i. We will continue to use $\oplus$ to represent the updated belief but it will be convenient not to use a function notation. We assume that both β and α are a set of disjunctions in minimal disjunctive form. As we will next show, there is no obvious unique way to define $\oplus$ in this context. There are a number of options of which we will consider two.

EXAMPLE 6. Let $\Phi_E = \{p, q, r\}$, $\beta = \{p, q, r\}$, $\alpha = \{\neg p \vee \neg q \vee \neg r\}$. The agent believes that all environment variables are true. The principal then announces that at least one is false, but not which one(s). Clearly, α contradicts β. The question is how the agent should update its belief. In one approach, that we call the optimistic update, the agent tries to keep as much of its belief as possible while avoiding a contradiction. One could, in distinction to the optimistic update, retain two of the atoms, say p and q, and change r to $\neg r$. The problem is that there is no unique way to make the choice of the two atoms to keep, as the agent could have kept p and r, for instance. Instead, we make new disjunctions that come as close as possible to the original belief without causing a contradiction, noting that α allows for any two of the atoms to remain true. The optimistic update in this case is $\{p \vee q, p \vee r, q \vee r, \neg p \vee \neg q \vee \neg r\}$. The other approach that we call the cautious approach deletes all beliefs that were involved in a contradiction with α. The cautious approach in this requires deleting each atom, yielding $\{\neg p \vee \neg q \vee \neg r\}$.

EXAMPLE 7. Let $\Phi_e = \{p, q, r\}$, $\beta = \{p, q, r\}$, $\alpha = \{\neg p \vee \neg q\}$. This is similar to the previous example, except that α says nothing about r. In this case the optimistic update is $\{p \vee q, r, \neg p \vee \neg q\}$, while the cautious update is $\{r, \neg p \vee \neg q\}$.

The Optimistic Belief Update: We now define the optimistic update in several steps. We use the notation $\oplus_o$ for this concept. We start with the environment variables Φ_E, the agent's belief β, and the announcement α, where β and α are in minimal disjunctive form. There are two cases. First, if $\beta \cup \alpha$ is consistent then we define $\beta \oplus_o \alpha = md(\beta \cup \alpha)$. Next, suppose that $\beta \cup \alpha$ is inconsistent. Accepting α as true we will try to preserve as much of β as possible. Let $\beta_1, \ldots, \beta_k$ be all subsets of β such that each $\beta_i \cup \alpha$ is inconsistent and there is no proper subset of β_i whose union with α is inconsistent. We write $\beta_0 = \beta \setminus \cup_{i=1}^{k} \beta_i$. Each β_i may contain several disjunctions, say $\beta_i = \{d_{i1}, \ldots, d_{ik}\}$. Form all disjunctions from pairs of d_{ij}. For example, if $d_{i_1} = p \vee q$ and $d_{i_2} = r$ then by $d_{i_1} \vee d_{i_2}$ we mean $p \vee q \vee r$. Let $\beta'_i = \{d_{i_m} \vee d_{i_n} | 1 \leq m < n \leq k\}$. Note that if $|\beta_i| = 1$ then $\beta'_i = \emptyset$. $\beta_i \cup \alpha$ must be consistent: otherwise there would be some proper subset of β_i whose union with α is inconsistent. We define $\beta \oplus_o \alpha = md(\beta_0 \cup \cup_{i=1}^{k} \beta'_i \cup \alpha)$.

Let us now show that $\oplus_o$ conforms with the $\oplus$ we defined previously for simple conjunctive beliefs and announcements. However, instead of functions, we write both β and α in simple disjunctive form. Without loss of generality assume that β contains all atoms and α may contain negations of atoms: $\beta = \{p_1, \ldots, p_n\}$, $\alpha = \{\ell_1, \ldots, \ell_m\}$. If $\beta \cup \alpha$ is consistent, then $\beta \oplus_o \alpha = \{p_1, \ldots, p_n, \ell_1, \ldots, \ell_m\}$ with repetitions removed. In case $\beta \cup \alpha$ is inconsistent, without loss of generality, let $\alpha = \{\neg p_1, \ldots, \neg p_i, \ell_{i+1}, \ldots, \ell_m\}$, that is, the inconsistency is caused by the incorrect beliefs for the variables $p_1, \ldots, p_i$. According to the above definition we form $\beta_1 = \{p_1\}, \ldots, \beta_i = \{p_i\}$. As $|\beta_1| = \ldots |\beta_i| = 1$, $\beta'_1 = \cdots \beta'_i = \emptyset$. Then by the definition $\beta \oplus_o \alpha = \{p_{i+1}, \ldots, p_n, \ell_1, \ldots, \ell_m\}$ (with repetitions removed).

Commutativity of Belief Updates: We would like the new belief updates to also have another important property, namely, that in the case of several announcements the order does not matter, that is, $(\beta \oplus \alpha_1) \oplus \alpha_2 = (\beta \oplus \alpha_2) \oplus \alpha_1$. It is easy to show that this property holds in the simple conjunctive case. Our next examples show why the order matters for $\oplus_o$.

Example 8. $\Phi_E = \{p, q\}$, $\beta = \{p\}$, $\alpha_1 = \{\neg p \vee q\}$, $\alpha_2 = \{\neg p\}$. Then $\beta \oplus_o \alpha_1 = \{p, q\}$ (there was no inconsistency), so $(\beta \oplus_o \alpha_1) \oplus_o \alpha_2 = \{\neg p, q\}$ (p was inconsistent with $\neg p$ and hence deleted). But $\beta \oplus_o \alpha_2 = \{\neg p\}$ and then $(\beta \oplus_o \alpha_2) \oplus_o \alpha_1 = \{\neg p\}$ (the result is placed in minimal disjunctive form). Hence $(\beta \oplus_o \alpha_1) \oplus_o \alpha_2 \neq (\beta \oplus_o \alpha_2) \oplus_o \alpha_1$.

In this example the problem was that the first announcement led the agent to conclude q. But the second announcement subsumes the first and has no q in it. By announcing α_2 first, α_1 becomes superfluous. The next example shows that even if there is no subsumption between announcements, as long as the announcements share a variable the problem of noncommutativity persists.

EXAMPLE 9. $\Phi_E = \{p, q, r\}$, $\beta = \{p, q \vee r\}$, $\alpha_1 = \{\neg p \vee \neg q\}$, $\alpha_2 = \{\neg p \vee \neg r\}$. Then $\beta \oplus_o \alpha_1 = \{p, \neg q, r\}$, and $(\beta \oplus_o \alpha_1) \oplus_o \alpha_2 = \{p \vee r, \neg q, \neg p \vee \neg r\}$, while $\beta \oplus_o \alpha_2 = \{p, \neg r, q\}$, and $(\beta \oplus_o \alpha_2) \oplus_o \alpha_1 = \{p \vee q, \neg r, \neg p \vee \neg q\}$. Again, $(\beta \oplus_o \alpha_1) \oplus_o \alpha_2 \neq (\beta \oplus_o \alpha_2) \oplus_o \alpha_1$.

We can show that if the announcements refer to different variables, we get commutativity. Previously we wrote $vars(\varphi)$ to refer to the variables of a formula. We extend this definition so that for a set of formulae S we write $vars(S) = \{vars(\varphi) \mid \varphi \in S\}$.

PROPOSITION 7. *If* $vars(\alpha_1) \cap vars(\alpha_2) = \emptyset$ *then* $(\beta \oplus \alpha_1) \oplus \alpha_2 = (\beta \oplus \alpha_2) \oplus \alpha_1$.

PROOF. Let $F = \{\varphi \in \beta \mid vars(\varphi) \cap vars(\alpha_1) \neq \emptyset$ and $vars(\varphi) \cap vars(\alpha_2) \neq \emptyset\}$. If $F = \emptyset$ then the interaction of β with α_1 must be different from the interaction of β with α_2. Hence the announcements must change different formulae of β (if any) and each formula in β can be inconsistent only with formulae entirely in α_1 or α_2. Therefore the order of application of α_1 and α_2 does not matter.

If $F \neq \emptyset$ let $\varphi \in \beta$ such that some variable of φ occurs in a formula of α_1 and some variable of φ occurs in a formula of α_2. The situation will be something like the following: $\varphi = p \vee q \vee$ (other literals), α_1 has a formula ψ_1 containing p or $\neg p$ and α_2 has a formula ψ_2 containing q or $\neg q$. Two types of interactions are possible with φ. The first type is where either ψ_1 or ψ_2 (or both) subsumes φ. In this case in both update orders φ is replaced by ψ_1 and ψ_2. The second type is where the interaction is caused by a negated literal in ψ_1 or ψ_2 (or both), such as if $\psi_1 = \neg p$. Then, when α_1 is applied, φ becomes $\varphi' = q \vee$ (other literals), but the update order makes no difference. ■

The Cautious Belief Update: We just sketch here the definition of the cautious belief update and some results about it based on the presentation of the optimistic belief update. There is only one place in the definition where $\oplus_c$ differs from $\oplus_o$. Recall that $\beta_1, \ldots, \beta_k$ are all the subsets of β that are minimally inconsistent with α and $\beta_0 = \beta \setminus \cup_{i=1}^{k} \beta_i$. For the optimistic update we tried to preserve as much of the information of β as possible by taking pairwise disjunctions. The idea of the cautious update is that we no longer trust any belief formula that is involved in an inconsistency with the announcement. Hence we define $\beta \oplus_c \alpha = md(\beta_0 \cup \alpha)$.

It is clear that $\oplus_c$ works the same way as $\oplus_o$ for simple conjunctive beliefs and announcements. Consider now what happens if the order of two announcements is switched. Examples 8 and 9 again show noncommutativity. In Example 8 the results are the same as before. In Example 9 $(\beta \oplus_0 \alpha_1) \oplus_0 \alpha_2 = \{\neg q, \neg p \vee \neg r\}$, while $(\beta \oplus_0 \alpha_2) \oplus_0 \alpha_1 = \{\neg r, \neg p \vee \neg q\}$. Also, Proposition 7 is proved the same way for $\oplus_c$.

Stability: Recall that our goal for announcements was to allow the principal to create stability in an unstable situation. The principal's strategy is to make announcements, possibly different ones to different agents, convincing them that their goal is not really dependent on the actions of other agents. A reasonable question to ask is whether moving to complex announcements will allow the principal to stabilise an unstable system that it cannot stabilise by simple conjunctive announcements. We present an example based on Example 3, to show that this is indeed the case.

Example 10. Consider a game with:

- $Ag = \{1, 2, 3\}$;
- $\Phi = \{p_1, p_2, p_3, q_1, q_2, q_3, q_4\}$,
- $\Phi_i = \{p_i\}$ (hence the q's are the environment variables);
- $\gamma_1 = p_1 \vee p_2 \vee q_1$;
- $\gamma_2 = p_2 \vee p_3 \vee q_2$;
- $\gamma_3 = p_3 \vee p_1 \vee q_3$;
- $\beta_1 = \{q_4\}$;
- $\beta_2 = \{\neg q_2\}$;
- $\beta_3 = \{\neg q_3\}$;
- $c(p_i, \top) = 1$ for all i and
- $v_E(q_i) = \bot$.

The difference between this example and Example 3 is slight, including the fact that in this example each environment variable has truth value $\bot$. There is also an extra variable, q_4, and the agents' beliefs do not involve all the environment variables. Exactly as in Example 3 the system is unstable. In that example the principal was able to stabilise the system by announcing either $\{q_1\}$ or $\{q_2\}$, or $\{q_3\}$ breaking the dependency cycle and making the irreflexive portion of the digraph acyclic. That is not possible to do here because these environment variables are false. But if the principal announces the true disjunctive formula $\alpha_1 = q_1 \vee \neg q_4$ to agent 1, notice what happens: $\beta_1 \oplus_x \alpha_1 = \{q_4, q_1\}$ (we wrote the subscript x because both versions of $\oplus$ give the same result). Now, the presence of q_1 in the belief of agent 1 breaks the cycle and the system becomes stable, even though no simple conjunctive announcement (that is true) can make the system stable.

In this example the system was stabilised by the principal essentially misleading an agent. While a truthful principal, restricted to simple conjunctive announcements can hide information from an agent, it cannot mislead an agent.

4.2. Beliefs and Possible Worlds

So far in this paper, we have considered models of belief that are essentially *syntactic* in nature. That is, beliefs can be understood as sets of formulae, (or in the simplest case, just one formula), which represents how that agent sees the environment variables. An extremely powerful alternative model involves a *possible worlds* model of beliefs [10]. In this approach, we characterise an agent's beliefs as a set of alternatives, called *possible worlds*, each one representing one possible way the environment variables could be, given the agent's beliefs. An agent is then said to believe a proposition if that proposition is true in all that agent's possible worlds. To adopt this approach for our scheme, first recall that $\mathcal{V}_E$ is the set of possible valuations for the environment variables Φ_E. Then, the beliefs of agent i are given as a subset $\beta_i \subseteq \mathcal{V}_E$, and we assume that $v_E \in \beta_i$ (i.e., each agent i considers the *actual* valuation of the environment variables to be possible). Where $\varphi \in \mathcal{L}$, we will write $B_i\varphi$ ("agent i believes φ") to mean that $v \models \varphi$ for all $v \in \beta_i$.

Given this model, we can define updates of beliefs with respect to arbitrary formulae, as follows. Where an announcement α is a formula of propositional logic over the variables Φ_E, we denote by $\beta_i \oplus \alpha$ the set:

$$\beta_i \oplus \alpha = \{v \in \beta_i \mid v \models \alpha\}.$$

Thus, in computing the update $\beta_i \oplus \alpha$ in this case, we simply eliminate from the set β_i of epistemic alternatives for agent i all possibilities that are not consistent with the new information α. Notice that since we require announcements to be truthful (i.e., any announcement α must satisfy the requirement that $v_E \models \alpha$), and that $v_E \in \beta_i$, then the update $\beta_i \oplus \alpha$ will always be well defined and the result of the update will always be non-empty.

The great advantage of this approach is that it allows us to capture *uncertainty* about the state of environment variables. Agent i may have $\beta_i = \{v^1, v^2\}$, with $v^1(p) = \top$ and $v^2(p) = \bot$ and meaning that i believes it is possible that p is true, and also believes it possible that p is false. Furthermore, i may have $v^1(q) = \bot$ and $v^2(q) = \top$. Thus we have $B_i(p \vee q)$ but not $B_i p$ or $B_i q$.

The key difficulty with this approach with respect to our basic model is how to define utility: our original formulation of (subjective) utility required that an agent had a definite opinion about all environment variables in order to compute subjective utility, whereas in the new model, an agent can be uncertain about the values of environment variables. How should such uncertainty be reflected in an agent's subjective assessment of utility? There seems to be no clearcut answer to this question, but the simplest approach might be a *pessimistic* view: the utility an agent i believes it will get from an outcome $(v_1, \ldots, v_n)$ is the utility it would get if the worst case possibility were true, according to its beliefs $\beta_i \subseteq \mathcal{V}_E$. We define *pessimistic subjective utility* through a function $\hat{u}_i(\cdots)$ as follows:

$$\hat{u}_i(v_1, \ldots, v_n) = \min\{u_i(v_1, \ldots, v_n, v') \mid v' \in \beta_i\}.$$

We can then define Nash stability with respect to the utility functions $\hat{u}_i$, and the same basic approach discussed above can then be applied. Note, however, that with this approach, it is possible for the principal to make arbitrary truthful statements about the values of the environment variables. For example, if there are two environment variables, p and q, both of which are in fact true, then the principal could announce $p \vee q$, thereby revealing that one of them is true without revealing that both are true.

5. Related Work

In the sense that the main thrust of our work is to design announcements that will modify games in such a way that certain outcomes are achieved in equilibrium, our work is similar in spirit to mechanism design/implementation theory, where the goal is to design games in such a way that certain outcomes are achieved in equilibrium [18]. However, we are aware of no work within the AI/computer science community that addresses the problem of manipulating games in the same way as we do — through communication.

Similar ideas to our dependency graph are used in the analysis of games in, e.g., [12, 5]. For example, Gottlob *et al.* investigate cases where it is computationally easy to compute pure strategy Nash equilibria in games; their main results relate to games in which the "neighbourhood graph" (which is essentially our dependency graph) is "small". They show that while this condition is not sufficient in itself to guarantee tractability, by combining it with a further condition, tractability can be obtained.

Work that has considered manipulating games within the AI/computer science community has focussed on the design of taxation schemes to influence behaviour [19, 2, 9]. For example, Endriss *et al.* consider the possibility of overlaying Boolean games with taxation schemes so that, if every player acts rationally, then a certain objective, represented as a Boolean formula Υ, will be satisfied in some Nash equilibrium of the resulting system [9].

Our work is also about the effect of making announcements, and in this sense it has some affinity with the growing body of work on *dynamic epistemic logic* (DEL) [7]. DEL tries to give a logical account of how the knowledge states of agents in a system are affected by announcements that take the form of logical formulae. Of particular interest in DEL are announcements that themselves refer to the knowledge of participants, which can affect systems in subtle and complex ways.

Also relevant is the substantial body of work on *speech acts*. Theories of speech acts are pragmatic theories of language: theories concerning the way that language is used. The theory of speech acts is usually seen as originating in the work of John Austin in the 1960s. In his seminal book *How to Do Things With Words*, he observed that certain types of natural language utterance *change the state of the world* [1]. To take a paradigm example, if a priest or other legally empowered individual utters the sentence "I now pronounce you man and wife" in the appropriate circumstances, then after the utterance, the legal relationships between the individuals involved in the utterance will have changed. Other examples of "formal" speech acts include christening and declaring war. In this way, utterances can be understood as changing the world in ways beyond their immediate physical effects (such as shouting and causing an avalanche). Now, a fundamental tenet of speech act theory is that utterances are actions, which are made by rational agents in the furtherance of their goals and preferences. And if they are actions, then this suggests that formalisms developed for reasoning about actions can be applied to reasoning about speech acts; such formalisms include Floyd-Hoare logic [17], dynamic logic [15], and the STRIPS formalism [11]. This observation led researchers to apply formalisms developed for reasoning about actions to the formalisation of speech acts; for example, Cohen and Perrault [6] used a STRIPS-style notation to formalise speech acts, where the pre- and post-conditions of speech acts are formulated with respect to the beliefs and desires of the participants in the communicative act. It would be intriguing to investigate the links between speech acts and our work: our announcements are, after all, nothing more than declarative speech acts. More generally, we could see speech acts as actions performed by players in a game, with the intention of modifying the resulting behaviour of other players in the game.

6. Conclusions

We have considered the general problem of manipulating games through communication: by making announcements in a game, we change the beliefs of the players of the game, and in this way we can perturb their choices. We have focussed mainly on the idea of stabilising games.

There are *many* obvious avenues for future research. We might consider richer models of belief (probabilistic and Bayesian models [14]), and of course, mixed strategy equilibria. We might consider the possibility of the principal lying, and of noisy communication. We might also consider announcements that refer to the epistemic state of agents ("player one knows the value of x"); this would take us close to the realm of dynamic epistemic logic [7].

Acknowledgements. An earlier version of this paper was presented at the Twenty Second International Joint Conference on Artificial Intelligence (IJCAI-2011 – Barcelona, Catalonia, July 2011). Many thanks to the anonymous IJCAI-2011 reviewer who pointed us at [4], the results of which helped us simplify our presentation considerably. We are grateful to the Studia Logica reviewer who suggested the concept of an announcement that leads to a unique equilibrium along with an example. This research was financially supported by the UK Royal Society, MOST (#3-6797), ISF (#1685). Kraus and Wooldridge gratefully acknowledge the support of the European Research Council under grants 267523 and 291528 respectively.

References

[1] Austin, J. L., *How to Do Things With Words*. Oxford University Press: Oxford, England, 1962.

[2] Bachrach, Y., E. Elkind, R. Meir, D. Pasechnik, M. Zuckerman, J. Rothe, and J. S. Rosenschein. The cost of stability in coalitional games. In *Proceedings SAGT 2009*, 2009.

[3] Bonzon, E., M.-C. Lagasquie, J. Lang, and B. Zanuttini. Boolean games revisited. In *Proceedings of the Seventeenth European Conference on Artificial Intelligence (ECAI-2006)*, Riva del Garda, Italy, 2006.

[4] Bonzon, E., M.-C. Lagasquie-Schiex, and J. Lang. Dependencies between players in Boolean games. *International Journal of Approximate Reasoning* 50:899–914, 2009.

[5] Brafman, R., C. Domshlak, Y. Engel, and M. Tennenholtz. Planning games. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*, 2009.

[6] Cohen, P. R., and C. R. Perrault. Elements of a plan based theory of speech acts. *Cognitive Science* 3:177–212, 1979.

[7] van Ditmarsch, H., W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer-Verlag: Berlin, Germany, 2007.

[8] Dunne, P. E., S. Kraus, W. van der Hoek, and M. Wooldridge. Cooperative boolean games. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2008)*, Estoril, Portugal, 2008.

[9] Endriss, U., S. Kraus, J. Lang, and M. Wooldridge. Designing incentives for boolean games. In *Proceedings of the Tenth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2011)*, Taipei, Taiwan, 2011.

[10] Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. The MIT Press: Cambridge, MA, 1995.

[11] Fikes, R. E., and N. Nilsson. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2:189–208, 1971.

[12] Gottlob, G., G. Greco, and F. Scarcello. Pure nash equilibria: Hard and easy games. *Journal of AI Research* 24:357–406, 2005.

[13] Grant, J., S. Kraus, and M. Wooldridge. Intentions in equilibrium. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-2010)*, Atlanta, GA, 2010.

[14] Halpern, J., *Reasoning about Uncertainty*. The MIT Press: Cambridge, MA, 2003.

[15] Harel, D., D. Kozen, and J. Tiuryn. *Dynamic Logic*. The MIT Press: Cambridge, MA, 2000.

[16] Harrenstein, P., W. van der Hoek, J.-J.Ch. Meyer, and C. Witteveen. Boolean games. In J. van Benthem, (ed.), *Proceeding of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge (TARK VIII)*, Siena, Italy, 2001, pp. 287–298.

[17] Hoare, C. A. R., An axiomatic basis for computer programming. *Communications of the ACM* 12(10):576–583, 1969.

[18] Maskin, E., The theory of implementation in Nash equilibrium: A survey. MIT Department of Economics Working Paper, 1983.

[19] Monderer, D., and M. Tennenholtz. K-implementation. *Journal of AI Research* 21:37–62, 2004.

[20] Osborne, M. J., and A. Rubinstein. *A Course in Game Theory*. The MIT Press: Cambridge, MA, 1994.

[21] Papadimitriou, C. H., *Computational Complexity*. Addison-Wesley: Reading, MA, 1994.

John Grant
Department of Mathematics
Towson University, Towson, MD 21252, USA
jgrant@towson.edu

Sarit Kraus
Department of Computer Science
Bar-Ilan University, Ramat-Gan, 52900 Israel
sarit@cs.biu.ac.il

Michael Wooldridge
Department of Computer Science
University of Oxford, Oxford OX1 3QD, UK
mjw@cs.ox.ac.uk

Inon Zuckerman
Industrial Engineering and Management Department
Ariel University, 40700 Israel
inonzu@ariel.ac.il