Nash Equilibria in Multi-agent Deontic Logic

Allard Tamminga

Department of Theoretical Philosophy
University of Groningen
The Netherlands

SOCREAL 2010
Hokkaido University
Sapporo, 27 March 2009

Outline

- Extrinsic Preferences
- Act Consequentialism
- Language of Multi-agent Deontic Logic
- Consequentialist Models
- Absolute and Conditional $\mathcal{F}$-Dominance
- Semantics of Multi-agent Deontic Logic
- Strategic Games and Nash Equilibria
- From Strategic Games to Consequentialist Frames
- A Formal Characterization of Nash Equilibria

Extrinsic Preferences

How to bring deontic logic and game theory together conceptually?

- It is perfectly possible that I have the obligation to do X, but at the same time prefer not to do X.
- We make a distinction between *extrinsic* preferences (which are the result of a previous judgment of betterness on the basis of reasons) and *intrinsic* preferences (which reflect the unreasoned subjective likings of the agents concerned).
- It makes perfect sense that I have the obligation to do X and at the same time *intrinsically* prefer not to do X (I just don't feel like it). Some intellectual effort is needed, however, to imagine a situation where I have the obligation to do X and at the same time *extrinsically* prefer not to do X.
- Hence, we assume that the preferences that figure in our deontic logic, guiding agents in evaluating the moral rightness of their actions, are extrinsic.

Act Consequentialism

- It still might be that I have the obligation to do X (I promised to do so) and at the same time extrinsically prefer not to do X.
- This possibility is minimized once we adopt an evaluative version of *act consequentialism* as our moral theory. In evaluative act consequentialism, the moral rightness of an action only depends on the value of its consequences.
- In evaluative act consequentialism using extrinsic preferences it is plausible that an action is obligatory if and only if that action is extrinsically preferred.

Language of Multi-agent Deontic Logic

Within a formal framework inspired by evaluative act consequentialism using extrinsic preferences, we study the logical behavior of two types of action permissions:

- *Absolute action permissions*: "In group $\mathcal{F}$'s interest, group $\mathcal{G}$ may perform action $\alpha_{\mathcal{G}}$" (abbreviated as $\mathsf{P}^{\mathcal{F}}_{\mathcal{G}}\alpha_{\mathcal{G}}$)
- *Conditional action permissions*: "If group $\mathcal{H}$ were to perform action $\alpha_{\mathcal{H}}$, then, in group $\mathcal{F}$'s interest, group $\mathcal{G}$ may perform action $\alpha_{\mathcal{G}}$" (abbreviated as $\mathsf{P}^{\mathcal{F}}_{\mathcal{G}}(\alpha_{\mathcal{G}}/\alpha_{\mathcal{H}})$).

Language of Multi-agent Deontic Logic

$\mathfrak{L}$ is a propositional modal language built from a countable set $\mathfrak{A} = \{\alpha_{\mathcal{G}}^{n} : \mathcal{G} \subseteq \mathcal{N} \text{ and } n \in \mathbb{N}\}$ of atomic propositions.

$\mathfrak{L}$ is the smallest set satisfying the following conditions:

- $\mathfrak{A} \subseteq \mathfrak{L}$
- If $\varphi \in \mathfrak{L}$, then $\neg\varphi \in \mathfrak{L}$
- If $\varphi, \psi \in \mathfrak{L}$, then $\varphi \wedge \psi \in \mathfrak{L}$
- If $\alpha_{\mathcal{G}} \in \mathfrak{A}$ and $\mathcal{F} \subseteq \mathcal{N}$, then $\mathsf{P}_{\mathcal{G}}^{\mathcal{F}} \alpha_{\mathcal{G}} \in \mathfrak{L}$
- If $\alpha_{\mathcal{G}}, \alpha_{\mathcal{H}} \in \mathfrak{A}$ and $\mathcal{F} \subseteq \mathcal{N}$ and $\mathcal{H} \subseteq \mathcal{N} - \mathcal{G}$, then $\mathsf{P}_{\mathcal{G}}^{\mathcal{F}}(\alpha_{\mathcal{G}}/\alpha_{\mathcal{H}}) \in \mathfrak{L}$.

Language of Multi-agent Deontic Logic

Two examples:

- $\mathsf{P}_i^i \alpha_i$ means "In his own interest, agent i may perform action α_i".
- $\mathsf{P}_{\mathcal{G}}^{\mathcal{N}}(\alpha_{\mathcal{G}}/\alpha_i)$ means "If agent i were to perform action α_i, then, in the grand coalition's interest, group $\mathcal{G}$ may perform action $\alpha_{\mathcal{G}}$".

Consequentialist Frames 1

A *consequentialist frame* $\mathfrak{F}$ is a quadruple $\langle \mathcal{W}, \mathcal{N}, Choice, (\succeq_{\mathcal{F}}) \rangle$, where $\mathcal{W}$ is a non-empty set of possible worlds, $\mathcal{N}$ is a finite set of agents, *Choice* is a choice function, and $\succeq_{\mathcal{F}}$ is a reflexive, transitive, and complete relation on $\mathcal{W}$ for each $\mathcal{F} \subseteq \mathcal{N}$.

Choice sets of *individual agents* are given by a function $Choice : \mathcal{N} \to \mathscr{P}(\mathscr{P}(\mathcal{W}))$, such that

1. for each agent $i \in \mathcal{N}$ it holds that $Choice(i)$ is a partition of $\mathcal{W}$,
2. for each selection function s assigning to each agent $i \in \mathcal{N}$ a set of possible worlds $s(i)$ such that $s(i) \in Choice(i)$ it holds that $\bigcap_{i \in \mathcal{N}} s(i)$ is non-empty (*independence of agents*).

Consequentialist Frames 2

This choice function for individual agents is extended to a function $\mathit{Choice} : \mathscr{P}(\mathcal{N}) \to \mathscr{P}(\mathscr{P}(\mathcal{W}))$ for groups of agents. Let *Select* be the set of all selection functions s assigning to each individual agent $i \in \mathcal{N}$ an option $s(i) \in \mathit{Choice}(i)$. Then

$$\mathit{Choice}(\mathcal{G}) = \{\bigcap_{i \in \mathcal{G}} s(i) : s \in \mathit{Select}\},$$

if $\mathcal{G}$ is non-empty. Otherwise, $\mathit{Choice}(\mathcal{G}) = \{\mathcal{W}\}$.

Consequentialist Models

A *consequentialist model* $\mathfrak{M}$ is an ordered pair $\langle \mathfrak{F}, v \rangle$, where $\mathfrak{F}$ is a consequentialist frame and v an interpretation function that assigns to each atomic proposition $\alpha_{\mathcal{G}} \in \mathfrak{A}$ an action $v(\alpha_{\mathcal{G}}) \in Choice(\mathcal{G})$.

In a consequentialist model, each possible action of each group of agents is assumed to have a name, that is, for each $\mathcal{G} \subseteq \mathcal{N}$ and each $K \in Choice(\mathcal{G})$ there is an atomic proposition $\alpha_{\mathcal{G}} \in \mathfrak{A}$ such that $v(\alpha_{\mathcal{G}}) = K$.

Absolute and Conditional $\mathcal{F}$-Dominance

Let $\mathfrak{F}$ be a consequentialist frame. Let $\mathcal{F}, \mathcal{G} \subseteq \mathcal{N}$ and $\mathcal{H} \subseteq \mathcal{N} - \mathcal{G}$. Let $K, K' \in Choice(\mathcal{G})$ and $L \in Choice(\mathcal{H})$. Then

$K \geq^{\mathcal{F}}_{\mathcal{G}} K'$	iff	for all $S \in Choice(\mathcal{N} - \mathcal{G})$ and for all $w, w' \in \mathcal{W}$ it holds that if $w \in K \cap S$ and $w' \in K' \cap S$, then $w \succeq_{\mathcal{F}} w'$.
$K \geq^{\mathcal{F}}_{(\mathcal{G}/\mathcal{H},L)} K'$	iff	for all $S \in Choice((\mathcal{N} - \mathcal{G}) - \mathcal{H})$ and for all $w, w' \in \mathcal{W}$ it holds that if $w \in K \cap L \cap S$ and $w' \in K' \cap L \cap S$, then $w \succeq_{\mathcal{F}} w'$.

Semantics of Multi-agent Deontic Logic

Let $\mathfrak{M} = \langle \mathfrak{F}, v \rangle$ be a consequentialist model. Let $\mathcal{F}, \mathcal{G} \subseteq \mathcal{N}$ and let $\mathcal{H} \subseteq \mathcal{N} - \mathcal{G}$. Let $w \in \mathcal{W}$ and let $\alpha_{\mathcal{G}}, \alpha_{\mathcal{H}} \in \mathfrak{A}$ and $\varphi, \psi \in \mathfrak{L}$. Then

$\mathfrak{M}, w \models \alpha_{\mathcal{G}}$	iff	$w \in v(\alpha_{\mathcal{G}})$
$\mathfrak{M}, w \models \neg\varphi$	iff	$\mathfrak{M}, w \not\models \varphi$
$\mathfrak{M}, w \models \varphi \wedge \psi$	iff	$\mathfrak{M}, w \models \varphi$ and $\mathfrak{M}, w \models \psi$
$\mathfrak{M}, w \models \mathrm{P}^{\mathcal{F}}_{\mathcal{G}} \alpha_{\mathcal{G}}$	iff	for all K' in $Choice(\mathcal{G})$ with $K' \neq v(\alpha_{\mathcal{G}})$ it holds that $v(\alpha_{\mathcal{G}}) \geq^{\mathcal{F}}_{\mathcal{G}} K'$
$\mathfrak{M}, w \models \mathrm{P}^{\mathcal{F}}_{\mathcal{G}} (\alpha_{\mathcal{G}} / \alpha_{\mathcal{H}})$	iff	for all K' in $Choice(\mathcal{G})$ with $K' \neq v(\alpha_{\mathcal{G}})$ it holds that $v(\alpha_{\mathcal{G}}) \geq^{\mathcal{F}}_{(\mathcal{G}/\mathcal{H}, v(\alpha_{\mathcal{H}}))} K'$.

Semantics

Theorem

Let $\mathfrak{M} = \langle \mathfrak{F}, v \rangle$ *be a consequentialist model. Let* $\mathcal{F}, \mathcal{G} \subseteq \mathcal{N}$ *and let* $\mathcal{H} \subseteq \mathcal{N} - \mathcal{G}$*. Let* $\alpha_{\mathcal{G}} \in \mathfrak{A}$*. Then the following statements are equivalent:*

- $\mathfrak{M} \models \mathsf{P}^{\mathcal{F}}_{\mathcal{G}} \alpha_{\mathcal{G}}$
- $\mathfrak{M} \models \mathsf{P}^{\mathcal{F}}_{\mathcal{G}} (\alpha_{\mathcal{G}} / \alpha_{\mathcal{H}})$ *for all* $\alpha_{\mathcal{H}} \in \mathfrak{A}$

Strategic Games and Nash Equilibria

A *strategic game* is a triple $G = \langle N, (A_i), (\succsim_i) \rangle$, where N is a finite set of players, for each player $i \in N$ it holds that A_i is a non-empty set of actions available to player i, and for each player $i \in N$ it holds that $\succsim_i$ is a preference relation on the set of outcomes $A = \times_{i \in N} A_i$.

Preference relations $\succsim_i$ are assumed to be reflexive, transitive, and complete.

An outcome $a^* \in A$ is a *Nash equilibrium* of a strategic game $G = \langle N, (A_i), (\succsim_i) \rangle$ if and only if for each player $i \in N$ it holds that

$$(a^*_{-i}, a^*_i) \succsim_i (a^*_{-i}, a_i) \text{ for all } a_i \in A_i.$$

From Strategic Games to Consequentialist Frames

Let $G = \langle N, (A_i), (\succsim_i) \rangle$ be a strategic game. The quadruple $\mathfrak{T}(G) = \langle \mathcal{W}, \mathcal{N}, Choice, (\succeq_{\mathcal{F}}) \rangle$ is defined as follows:

- $\mathcal{W} = A$
- $\mathcal{N} = N$
- $Choice(\mathcal{G}) =$
$$\begin{cases} \{\{(a_{\mathcal{G}}, a_{-\mathcal{G}}) \in A : a_{-\mathcal{G}} \in A_{-\mathcal{G}}\} : a_{\mathcal{G}} \in A_{\mathcal{G}}\}, & \text{if } \mathcal{G} \neq \emptyset \\ \{\mathcal{W}\}, & \text{otherwise} \end{cases}$$
- $\succeq_{\mathcal{F}} = \begin{cases} \succsim_i, & \text{if } \mathcal{F} = \{i\} \\ \mathcal{W} \times \mathcal{W}, & \text{otherwise} \end{cases}$

Theorem

Let G be a strategic game. Then $\mathfrak{T}(G)$ is a consequentialist frame.

A Formal Characterization of Nash Equilibria

Finally, we define a suitable valuation function v_f that keeps track of which atomic proposition $\alpha_\mathcal{G}$ in $\mathfrak{A}_\mathcal{G}$ is validated by the performance of which action $a_\mathcal{G}$ in $A_\mathcal{G}$.

- Let f be an injective map that for each $\mathcal{G} \subseteq \mathcal{N}$ assigns to each action $a_\mathcal{G}$ in each $A_\mathcal{G}$ an atomic proposition $\alpha_\mathcal{G}$ in $\mathfrak{A}_\mathcal{G}$.
- If there is an action $a_\mathcal{G}$ in $A_\mathcal{G}$ such that $f(a_\mathcal{G}) = \alpha_\mathcal{G}$, we define $v_f(\alpha_\mathcal{G}) = \{(a_\mathcal{G}, a_{-\mathcal{G}}) \in A : a_{-\mathcal{G}} \in A_{-\mathcal{G}}\}$. If there is no action $a_\mathcal{G}$ in $A_\mathcal{G}$ such that $f(a_\mathcal{G}) = \alpha_\mathcal{G}$, then we simply put $v_f(\alpha_\mathcal{G}) = K$ for some unique designated $K \in \mathit{Choice}(\mathcal{G})$.

Theorem

Let G be a strategic game. Then

$$a^* \text{ is a Nash equilibrium of } G$$
$$\text{iff}$$
$$\langle \mathfrak{T}(G), v_f \rangle \models \bigwedge_{i \in \mathcal{N}} \mathsf{P}^i_i(f(a^*_i)/f(a^*_{-i})).$$