Conditional Obligations in Strategic Situations

NorMAS 2008, Université du Luxembourg, 16 July 2008

Barteld Kooi and Allard Tamminga*

Abstract

We investigate Horty's action constants and their functioning in action obligations of the form "In the interest of group $\mathcal{F}$ of agents, group $\mathcal{G}$ of agents ought to perform action K" and in conditional action obligations of the form "If group $\mathcal{H}$ of agents performs action L, then in the interest of group $\mathcal{F}$ of agents, group $\mathcal{G}$ of agents ought to perform action K". We show that a straightforward Horty-style analysis of action obligations leads to counterintuitive results. To make amends, we introduce weak action obligations. This new type of action obligations is exactly what we need to show that an action is obligatory for a group of agents if and only if this action is obligatory for this group of agents regardless of what the others do.

1 Moral Reasoning by Cases

> Consider a soldier at the front, waiting with his comrades to repulse an enemy attack. It may occur to him that if the defense is likely to be successful, then it isn't very probable that his own personal contribution will be essential. But if he stays, he runs the risk of being killed or wounded—apparently for no point. On the other hand, if the enemy is going to win the battle, then his chances of death or injury are higher still, and now quite clearly to no point, since the line will be overwhelmed anyway. Based on this reasoning, it would appear that the soldier is better off running away regardless of who is going to win the battle. (Ross, 2006)

Our soldier's line of reasoning can be rendered as follows:

*Department of Theoretical Philosophy, University of Groningen, Oude Boteringestraat 52, 9712 GL Groningen, E-mail: {B.P.Kooi, A.M.Tamminga}@rug.nl.

(1) “If we win the battle, then fleeing is the best way to further my self-interest.”
(2) “If we lose the battle, then fleeing is the best way to further my self-interest.”

Therefore,

(3) “Fleeing is the best way to further my self-interest.”

2 Language

We shall use a propositional modal language $\mathfrak{L}$ built from a countable set $\mathfrak{P} = \{p_1, p_2, \ldots\}$ of atomic propositions and a finite set $A = \{a_1, \ldots, a_n\}$ of individual agents. We employ $\mathcal{F}$ and $\mathcal{G}$ as variables for sets of agents. $\mathfrak{L}$ is given by the following Backus-Naur Form:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \odot^{\mathcal{F}}_{\mathcal{G}}\varphi \mid \odot^{\mathcal{F}}_{\mathcal{G}}(\varphi/\varphi).$$

3 Consequentialist Models

Definition 1 (Consequentialist Models) A *consequentialist model* $\mathfrak{M}$ is an ordered pair $\langle \mathfrak{S}, \mathfrak{I} \rangle$, where $\mathfrak{S}$ is a choice structure and $\mathfrak{I}$ an interpretation.

Definition 2 (Choice Structures) A *choice structure* $\mathfrak{S}$ is a triple $\langle W, A, \mathit{Choice} \rangle$, where W is a non-empty set of possible worlds, A a finite set of agents, and Choice a choice function.

Choice sets of *individual agents* are given by a function Choice from individual agents to sets of sets of possible worlds, meeting the conditions that (1) for each individual agent a in A it holds that $\mathit{Choice}(a)$ is a partition of W, and (2) for each selection function s assigning to each individual agent a in A a set of possible worlds $s(a)$ such that $s(a) \in \mathit{Choice}(a)$ it holds that $\bigcap_{a \in A} s(a)$ is non-empty.

Next, we extend this choice function for individual agents to a function Choice for *groups of agents*, given the set Select of selection functions s assigning to each individual agent a in A an option $s(a)$ in $\mathit{Choice}(a)$:

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

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

Definition 3 (Interpretations) An *interpretation* $\mathfrak{I}$ is an ordered pair $\langle \mathit{Utility}, V \rangle$, where $\mathit{Utility}$ is a utility function and V a valuation function.

The utility function $\mathit{Utility}$ is a function from ordered pairs consisting of a set of individual agents and a possible world to the real numbers between, say, -5 and 5. Thus, if a group $\mathcal{G}$ of agents assigns to a possible world w a utility of 4, we write $\mathit{Utility}(\mathcal{G}, w) = 4$.

The valuation function V is a function from ordered pairs consisting of an atomic propostion and a possible world to the truth-values TRUE and FALSE. Thus, if an atomic propostion p is true in a possible world w, we write $V(p, w) = \text{TRUE}$.

Definition 4 (Restricted Choice Sets) Let $\mathfrak{M}$ be a consequentialist model. Let $\mathcal{G} \subseteq A$ and $X \subseteq W$. Then

$$\mathit{Choice}(\mathcal{G}/X) \quad = \quad \{K \cap X \mid K \in \mathit{Choice}(\mathcal{G}) \text{ and } K \cap X \neq \emptyset\}.$$

Definition 5 ($\mathcal{F}$-dominance) Let $\mathfrak{M}$ be a consequentialist model. Let $\mathcal{F}, \mathcal{G} \subseteq A$ and $K, K' \in \mathit{Choice}(\mathcal{G})$ and $x \subseteq K$ and $x' \subseteq K'$. Then

$$x \succeq_{\mathcal{G}}^{\mathcal{F}} x' \quad \text{iff} \quad \text{for all } S \in \mathit{Choice}(A - \mathcal{G}) \text{ and for all } w, w' \in W \text{ it holds that if } w \in x \cap S \text{ and } w' \in x' \cap S \text{, then } \mathit{Utility}(\mathcal{F}, w) \geq \mathit{Utility}(\mathcal{F}, w').$$

Moreover, $x \succ_{\mathcal{G}}^{\mathcal{F}} x'$ if and only if $x \succeq_{\mathcal{G}}^{\mathcal{F}} x'$ and $x' \not\succeq_{\mathcal{G}}^{\mathcal{F}} x$.

4 Semantics

Definition 6 (Semantical Rules) Let $\mathfrak{M}$ be a consequentialist model. Let $w \in W$ and let $\varphi, \psi \in \mathfrak{L}$. Then

(i) $\mathfrak{M}, w \models p$ iff $V(p, w) = \text{TRUE}$, if $p \in \mathfrak{P}$

(ii) $\mathfrak{M}, w \models \neg\varphi$ iff $\mathfrak{M}, w \not\models \varphi$

(iii) $\mathfrak{M}, w \models \varphi \wedge \psi$ iff $\mathfrak{M}, w \models \varphi$ and $\mathfrak{M}, w \models \psi$

(iv) $\mathfrak{M}, w \models \odot_{\mathcal{G}}^{\mathcal{F}} \varphi$ iff for all K in $\mathit{Choice}(\mathcal{G})$ with $K \not\subseteq [\![\varphi]\!]_{\mathfrak{M}}$ there is a K' in $\mathit{Choice}(\mathcal{G})$ with $K' \subseteq [\![\varphi]\!]_{\mathfrak{M}}$ such that (1) $K' \succ_{\mathcal{G}}^{\mathcal{F}} K$, and (2) for all K'' in $\mathit{Choice}(\mathcal{G})$ with $K'' \succeq_{\mathcal{G}}^{\mathcal{F}} K'$ it holds that $K'' \subseteq [\![\varphi]\!]_{\mathfrak{M}}$.

(v) $\mathfrak{M}, w \models \odot_{\mathcal{G}}^{\mathcal{F}}(\psi/\varphi)$ iff for all x in $\mathit{Choice}(\mathcal{G}/\varphi)$ with $x \not\subseteq [\![\psi]\!]_{\mathfrak{M}}$ there is a x' in $\mathit{Choice}(\mathcal{G}/\varphi)$ with $x' \subseteq [\![\psi]\!]_{\mathfrak{M}}$ such that (1) $x' \succ_{\mathcal{G}}^{\mathcal{F}} x$, and (2) for all x'' in $\mathit{Choice}(\mathcal{G}/\varphi)$ with $x'' \succeq_{\mathcal{G}}^{\mathcal{F}} x'$ it holds that $x'' \subseteq [\![\psi]\!]_{\mathfrak{M}}$.

We use $\mathit{Choice}(\mathcal{G}/\varphi)$ as an abbreviation for $\mathit{Choice}(\mathcal{G}/[\![\varphi]\!]_{\mathfrak{M}})$.

4.1 The invalidity of our soldier's reasoning

We can now formally show that our soldier's reasoning is invalid:

Lemma 1 Let $\mathcal{F}, \mathcal{G} \subseteq A$. Let $\varphi, \psi \in \mathfrak{L}$. Then

$\not\models (\odot_{\mathcal{G}}^{\mathcal{F}}(\psi/\varphi) \wedge \odot_{\mathcal{G}}^{\mathcal{F}}(\psi/\neg\varphi)) \to \odot_{\mathcal{G}}^{\mathcal{F}}\psi.$

5 Action Constants and Action Obligations

Let us extend our language with a set $\mathfrak{A}$ of action constants. Given a choice structure $\mathfrak{S}$, we define the set $\mathfrak{A} = \{A_{\mathcal{G}}^K \mid \mathcal{G} \subseteq A \text{ and } K \in \mathit{Choice}(\mathcal{G})\}$ of action constants, which must be conceived as a subset of the set $\mathfrak{P}$ of atomic propositions. The valuation function for action constants is defined as follows:

Definition 7 (Action Constants) Let $\mathfrak{M}$ be a consequentialist model. Let $w \in W$ and let $A_{\mathcal{G}}^K \in \mathfrak{A}$. Then

$$V(A_{\mathcal{G}}^K, w) = \text{TRUE} \quad \text{iff} \quad w \in K.$$

From Definition 6 (iv) and (v) we obtain the following semantical rules for *strong action obligations* and for *strong conditional action obligations*:

Lemma 2 (Strong Action Obligations) Let $\mathfrak{M}$ be a consequentialist model and let $w \in W$. Let $\mathcal{G} \subseteq A$ and $\mathcal{H} \subseteq A - \mathcal{G}$. Let $A_{\mathcal{G}}^K, A_{\mathcal{H}}^L \in \mathfrak{A}$. Then

$\mathfrak{M}, w \models \odot_{\mathcal{G}}^{\mathcal{F}} A_{\mathcal{G}}^K$	iff	for all K' in $\mathit{Choice}(\mathcal{G})$ with $K' \neq K$ it holds that $K \succ_{\mathcal{G}}^{\mathcal{F}} K'$
$\mathfrak{M}, w \models \odot_{\mathcal{G}}^{\mathcal{F}}(A_{\mathcal{G}}^K/A_{\mathcal{H}}^L)$	iff	for all K' in $\mathit{Choice}(\mathcal{G})$ with $K' \neq K$ it holds that $K \cap L \succ_{\mathcal{G}}^{\mathcal{F}} K' \cap L$.

Lemma 3 Let $\mathcal{F}, \mathcal{G} \subseteq A$ and $\mathcal{H} \subseteq A - \mathcal{G}$. Let $A_{\mathcal{G}}^K, A_{\mathcal{H}}^L, A_{\mathcal{H}}^{L_1}, A_{\mathcal{H}}^{L_2} \in \mathfrak{A}$. Then

(i) $\models (\odot_{\mathcal{G}}^{\mathcal{F}}(A_{\mathcal{G}}^K/A_{\mathcal{H}}^{L_1}) \wedge \odot_{\mathcal{G}}^{\mathcal{F}}(A_{\mathcal{G}}^K/A_{\mathcal{H}}^{L_2})) \to \odot_{\mathcal{G}}^{\mathcal{F}}(A_{\mathcal{G}}^K/A_{\mathcal{H}}^{L_1} \vee A_{\mathcal{H}}^{L_2})$

(ii) $\models (\odot_{\mathcal{G}}^{\mathcal{F}}(A_{\mathcal{G}}^K/A_{\mathcal{H}}^L) \wedge \odot_{\mathcal{G}}^{\mathcal{F}}(A_{\mathcal{G}}^K/\neg A_{\mathcal{H}}^L)) \to \odot_{\mathcal{G}}^{\mathcal{F}} A_{\mathcal{G}}^K.$

Theorem 1 Let $\mathfrak{M}$ be a consequentialist model. Let $\mathcal{F}, \mathcal{G} \subseteq A$ and let $\mathcal{H} \subseteq A - \mathcal{G}$. Then (i) implies (ii):

(i) $\mathfrak{M} \models \odot_{\mathcal{G}}^{\mathcal{F}}(A_{\mathcal{G}}^K/A_{\mathcal{H}}^L)$ for all $L \in \mathit{Choice}(\mathcal{H})$

(ii) $\mathfrak{M} \models \odot_{\mathcal{G}}^{\mathcal{F}} A_{\mathcal{G}}^K.$

Just like in Horty's Proposition 5.14 (Horty 2001, p. 110), the converse of Theorem 1 does not hold. We make amends by a definition of weak action obligations:

Definition 8 (Weak Action Obligations) Let $\mathfrak{M}$ be a consequentialist model and let $w \in W$. Let $\mathcal{G} \subseteq A$ and $\mathcal{H} \subseteq A - \mathcal{G}$. Let $A^K_\mathcal{G}, A^L_\mathcal{H} \in \mathfrak{A}$. Then

$\mathfrak{M}, w \models \circledcirc^\mathcal{F}_\mathcal{G} A^K_\mathcal{G}$	iff	for all K' in $Choice(\mathcal{G})$ with $K' \neq K$ it holds that $K \succeq^\mathcal{F}_\mathcal{G} K'$
$\mathfrak{M}, w \models \circledcirc^\mathcal{F}_\mathcal{G} (A^K_\mathcal{G} / A^L_\mathcal{H})$	iff	for all K' in $Choice(\mathcal{G})$ with $K' \neq K$ it holds that $K \cap L \succeq^\mathcal{F}_\mathcal{G} K' \cap L$.

Theorem 2 Let $\mathfrak{M}$ be a consequentialist model. Let $\mathcal{F}, \mathcal{G} \subseteq A$ and let $\mathcal{H} \subseteq A - \mathcal{G}$. Then the following statements are equivalent:

(i) $\mathfrak{M} \models \circledcirc^\mathcal{F}_\mathcal{G} (A^K_\mathcal{G} / A^L_\mathcal{H})$ for all $L \in Choice(\mathcal{H})$
(ii) $\mathfrak{M} \models \circledcirc^\mathcal{F}_\mathcal{G} A^K_\mathcal{G}$.

6 Future Research

The present formalism allows for a definition of n-agent Nash-equilibria. For example, given a consequentialist model $\mathfrak{M}$ consisting of four worlds and two agents a and b, such that $Choice(a) = \{K_1, K_2\}$ and $Choice(b) = \{L_1, L_2\}$, it holds that $K \cap L$ is a Nash-equilibrium in $\mathfrak{M}$ if and only if $\mathfrak{M} \models \circledcirc^a_a (A^K_a / A^L_b) \wedge \circledcirc^b_b (A^L_b / A^K_a)$.

To systematically study these relations between game theory and our consequentialist multi-agent deontic logic, it would be appropriate to formalize conditional action obligations as $[A^L_\mathcal{H}] \circledcirc^\mathcal{F}_\mathcal{G} A^K_\mathcal{G}$ and to interpret them dynamically, that is, a conditional action obligation of the form $[A^L_\mathcal{H}] \circledcirc^\mathcal{F}_\mathcal{G} A^K_\mathcal{G}$ is true in a model $\mathfrak{M}_1$ if and only if the absolute action obligation $\circledcirc^\mathcal{F}_\mathcal{G} A^K_\mathcal{G}$ is true in the model $\mathfrak{M}_2$ that results from *updating* $\mathfrak{M}_1$ with the action $A^L_\mathcal{H}$. We conjecture that this new language for action obligations and its corresponding dynamic semantics also allows for a definition of subgame perfect equilibria.

References

Horty, J. F. (2001). *Agency and Deontic Logic*. New York: Oxford University Press.

Kooi, B. P. & A. M. Tamminga (2008). Moral conflicts between groups of agents. *Journal of Philosophical Logic*, 37, 1–21.

Ross, D. (2006). Game theory. In: E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy*. URL=<http://plato.stanford.edu/entries/game-theory>.