Moral Conflicts between Groups of Agents

Allard Tamminga*

SOCREAL 2007
Hokkaido University
March 10, 2007

1 Introduction

Two groups of agents, $\mathcal{G}_1$ and $\mathcal{G}_2$, face a *moral conflict* if $\mathcal{G}_1$ has a moral obligation and $\mathcal{G}_2$ has a moral obligation, such that these obligations cannot both be fulfilled.

> [B]efore we consider what actions are good or bad, right or wrong, it is proper to consider first what is meant by, and what not, (..) the expression 'doing an action' or 'doing something' (Austin 1957, p. 178)

In our analysis of moral conflicts between groups of agents, we adopt Austin's suggestion. On the basis of the well established *stit* logics of agency developed by Nuel Belnap and others,[1] we present a consequentialist system of multi-agent deontic logic, which is a generalization of John Horty's utilitarian deontic logic.[2]

$\Diamond\phi$	'It is possible that ϕ'
$[\mathcal{G}]\phi$	'Group $\mathcal{G}$ of agents sees to it that ϕ'
$\odot^{\mathcal{F}}_{\mathcal{G}}\phi$	'In the interest of group $\mathcal{F}$ of agents, group $\mathcal{G}$ of agents ought to see to it that ϕ'

Two groups of agents, $\mathcal{G}_1$ and $\mathcal{G}_2$, face a *basic moral conflict* if and only if there is a formula ϕ, such that both $\odot^{\mathcal{F}_1}_{\mathcal{G}_1}\phi$ and $\odot^{\mathcal{F}_2}_{\mathcal{G}_2}\neg\phi$ are true.

*Department of Theoretical Philosophy, Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands. E-mail: A.M.Tamminga@rug.nl. Joint work with Barteld Kooi.

[1] N. Belnap, M. Perloff & M. Xu (2001). *Facing the Future. Agents and Choices in Our Indeterminist World*. New York: Oxford University Press.

[2] J. Horty (2001). *Agency and Deontic Logic*. New York: Oxford University Press.

2 Language and Semantics

We use a 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. $\mathfrak{L}$ is the smallest set (in terms of set-theoretical inclusion) satisfying the conditions (i) through (v):

(i) $\mathfrak{P} \subseteq \mathfrak{L}$
(ii) If $\phi \in \mathfrak{L}$ and $\psi \in \mathfrak{L}$, then $(\phi \wedge \psi) \in \mathfrak{L}$ and $(\phi \rightarrow \psi) \in \mathfrak{L}$
(iii) If $\phi \in \mathfrak{L}$, then $\neg\phi \in \mathfrak{L}$ and $\Diamond\phi \in \mathfrak{L}$
(iv) If $\phi \in \mathfrak{L}$ and $\mathcal{G} \subseteq A$, then $[\mathcal{G}]\phi \in \mathfrak{L}$
(v) If $\phi \in \mathfrak{L}$ and $\mathcal{F} \subseteq A$ and $\mathcal{G} \subseteq A$, then $\odot_{\mathcal{G}}^{\mathcal{F}}\phi \in \mathfrak{L}$.

2.1 Consequentialist Models

Definition 1 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.

2.2 Choice Structures

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

2.2.1 Choice functions

Given a non-empty set W of possible worlds and a finite set A of individual agents, we define choice sets of individual agents by a choice function from individual agents to sets of sets of possible worlds, *i.e.*, $Choice : A \mapsto \wp(\wp(W))$, meeting the conditions that (1) for each individual agent a in A it holds that $Choice(a)$ 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 $s(a) \in Choice(a)$ it holds that $\bigcap_{a \in A} s(a)$ is non-empty.

Given a choice function $Choice$ from individual agents to sets of sets of possible worlds and given the corresponding set $Select$ of selection functions s assigning to each individual agent a in A an option $s(a)$ in $Choice(a)$, we define

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

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

2.2.2 $\mathcal{G}$-choice equivalence of worlds

In choosing an option K from $Choice(\mathcal{G})$, the group $\mathcal{G}$ of agents restricts the total set of possible worlds to the possible worlds in the set K. A formula of the form $[\mathcal{G}]\phi$, informally interpreted as 'Group $\mathcal{G}$ of agents sees to it that ϕ', is true in a world w if and only if ϕ is true in all possible worlds that are elements of the option of $\mathcal{G}$ that contains w. Or, equivalently, if and only if for all possible worlds w' that are *$\mathcal{G}$-choice equivalent* to world w it holds that ϕ is true in world w'.

Definition 3 ($\mathcal{G}$-Choice Equivalence) Let $\mathfrak{S}(= \langle W, A, Choice\rangle)$ be a choice structure. Let $\mathcal{G} \subseteq A$. Let $w, w' \in W$. Then $w \sim_{\mathcal{G}} w'$ (*w and w' are $\mathcal{G}$-choice equivalent*) is defined to be:

$$w \sim_{\mathcal{G}} w' \quad \text{iff} \quad \text{for all } K \in Choice(\mathcal{G}) \text{ with } w \in K \text{ it holds that } w' \in K.$$

2.3 Interpretations

Definition 4 An *interpretation* $\mathfrak{I}$ is an ordered pair $\langle Utility, V\rangle$, where $Utility$ is a utility function and V a valuation function.

2.3.1 Utility functions

We assume that individual utilities are given by a utility function from ordered pairs consisting of an individual agent and a possible world to the real numbers between -5 and 5, *i.e.*, $Utility : A \times W \mapsto [-5, 5]$.

The group utility a group $\mathcal{F}$ of agents assigns to a possible world w is defined as the arithmetical mean of the individual utilities the individual agents in $\mathcal{F}$ assign to w:

$$Utility(\mathcal{F}, w) = \frac{1}{|\mathcal{F}|} \sum_{a \in \mathcal{F}} Utility(a, w),$$

if $\mathcal{F}$ is non-empty. Otherwise, $Utility(\mathcal{F}, w) = 0$.

2.3.2 $\mathcal{F}$-dominance between $\mathcal{G}$'s options

Roughly, a formula of the form $\odot^{\mathcal{F}}_{\mathcal{G}}\phi$, informally interpreted as 'In the interest of group $\mathcal{F}$ of agents, group $\mathcal{G}$ of agents ought to see to it that ϕ', is true in a world w if and only if for all options K in $Choice(\mathcal{G})$ that do not ensure ϕ there is a strictly $\mathcal{F}$-better option K' in $Choice(\mathcal{G})$ such that (1) option K' ensures ϕ, and (2) all options K'' that are at least as $\mathcal{F}$-good as K' also ensure ϕ. We interpret "$\mathcal{F}$-betterness" decision-theoretically.

When a group $\mathcal{G}$ performs a collective action by choosing an option K from $Choice(\mathcal{G})$, it constrains the set W of possible worlds to that set K of possible worlds. It may be, however, that the agents who are not members of $\mathcal{G}$ (and who therefore are members of the group $A - \mathcal{G}$) perform a collective action by choosing an option S from $Choice(A - \mathcal{G})$, thereby constraining the set K to the set of possible worlds $K \cap S$. Hence, $\mathcal{G}$ usually will not be able to fully determine the outcome of its collective actions, since the final outcome also depends on the actions of agents in $A - \mathcal{G}$. Nevertheless, we can define an $\mathcal{F}$-dominance relation over $\mathcal{G}$'s options. If K and K' both are in $Choice(\mathcal{G})$, then, intuitively, K weakly $\mathcal{F}$-dominates K' if and only if option K promotes the utility of group $\mathcal{F}$ at least as well as option K', regardless of the collective action of the agents in $A - \mathcal{G}$.

Definition 5 ($\mathcal{F}$-Dominance) Let $\mathfrak{M}(= \langle \mathfrak{S}, \mathfrak{I} \rangle)$ be a consequentialist model. Let $\mathcal{F}, \mathcal{G} \subseteq A$ and let $K, K' \in Choice(\mathcal{G})$. Then $K \succeq_{\mathcal{G}}^{\mathcal{F}} K'$ (*K weakly $\mathcal{F}$-dominates K' for $\mathcal{G}$*) is defined to be:

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

As usual, $K \succ_{\mathcal{G}}^{\mathcal{F}} K'$ (*K strongly $\mathcal{F}$-dominates K' for $\mathcal{G}$*) if and only if $K \succeq_{\mathcal{G}}^{\mathcal{F}} K'$ and $K' \not\succeq_{\mathcal{G}}^{\mathcal{F}} K$.

2.4 Semantics

Definition 6 (Semantical Rules) Let $\mathfrak{M}(= \langle \mathfrak{S}, \mathfrak{I} \rangle)$ be a consequentialist model. Let $w \in W$ and let $\phi, \psi \in \mathfrak{L}$. Then

(i)	$\mathfrak{M}, w \models p$	iff	$V(p, w) = \text{TRUE}$,	if $p \in \mathfrak{P}$
(ii)	$\mathfrak{M}, w \models \neg\phi$	iff	$\mathfrak{M}, w \not\models \phi$	
(iii)	$\mathfrak{M}, w \models \phi \wedge \psi$	iff	$\mathfrak{M}, w \models \phi$ and $\mathfrak{M}, w \models \psi$	
(iv)	$\mathfrak{M}, w \models \phi \rightarrow \psi$	iff	$\mathfrak{M}, w \not\models \phi$ and/or $\mathfrak{M}, w \models \psi$	
(v)	$\mathfrak{M}, w \models \Diamond\phi$	iff	there is a w' in W such that $\mathfrak{M}, w' \models \phi$	
(vi)	$\mathfrak{M}, w \models [\mathcal{G}]\phi$	iff	for all w' in W with $w \sim_{\mathcal{G}} w'$ it holds that $\mathfrak{M}, w' \models \phi$	
(vii)	$\mathfrak{M}, w \models \odot_{\mathcal{G}}^{\mathcal{F}}\phi$	iff	for all K in $Choice(\mathcal{G})$ with $K \not\subseteq [\![\phi]\!]_{\mathfrak{M}}$ there is a K' in $Choice(\mathcal{G})$ with $K' \subseteq [\![\phi]\!]_{\mathfrak{M}}$ such that (1) $K' \succ_{\mathcal{G}}^{\mathcal{F}} K$, and (2) for all K'' in $Choice(\mathcal{G})$ with $K'' \succeq_{\mathcal{G}}^{\mathcal{F}} K'$ it holds that $K'' \subseteq [\![\phi]\!]_{\mathfrak{M}}$.	

We introduce the following notational conventions: Given a model $\mathfrak{M}$, we write $\mathfrak{M} \models \phi$, if for all worlds w in W it holds that $\mathfrak{M}, w \models \phi$. We write $\models \phi$, if for all models $\mathfrak{M}$ it holds that $\mathfrak{M} \models \phi$. Given a choice structure $\mathfrak{S}$, we write $\mathfrak{S} \models \phi$, if for all interpretations $\mathfrak{I}$ of $\mathfrak{S}$ it holds that $\langle \mathfrak{S}, \mathfrak{I} \rangle \models \phi$.

Lemma 1 Let $\phi, \psi \in \mathfrak{L}$. Then

(i) $\models \odot^{\mathcal{F}}_{\mathcal{G}} \phi \rightarrow \Diamond [\mathcal{G}] \phi$ (‘ought’ implies ‘can’)
(ii) If $\models \phi \leftrightarrow \psi$, then $\models \odot^{\mathcal{F}}_{\mathcal{G}} \phi \leftrightarrow \odot^{\mathcal{F}}_{\mathcal{G}} \psi$
(iii) If $\models \phi$, then $\models \odot^{\mathcal{F}}_{\mathcal{G}} \phi$
(iv) $\models \odot^{\mathcal{F}}_{\mathcal{G}} (\phi \wedge \psi) \rightarrow \odot^{\mathcal{F}}_{\mathcal{G}} \phi \wedge \odot^{\mathcal{F}}_{\mathcal{G}} \psi$
(v) $\models \odot^{\mathcal{F}}_{\mathcal{G}} \phi \wedge \odot^{\mathcal{F}}_{\mathcal{G}} \psi \rightarrow \odot^{\mathcal{F}}_{\mathcal{G}} (\phi \wedge \psi)$ (deontic agglomeration)

2.4.1 An example: the Prisoner’s Dilemma

	Don’t confess	*Confess*
Don’t confess	$3, 3$	$0, 4$
Confess	$4, 0$	$1, 1$

This payoff matrix can be translated into a consequentialist model $\mathfrak{M}(= \langle \mathfrak{S}, \mathfrak{I} \rangle)$. The choice structure $\mathfrak{S}$ is given by $W = \{w_1, w_2, w_3, w_4\}$, $A = \{a, b\}$, $Choice(a) = \{\{w_1, w_2\}, \{w_3, w_4\}\}$, and $Choice(b) = \{\{w_1, w_3\}, \{w_2, w_4\}\}$. The interpretation $\mathfrak{I}$ is given by

$$\begin{array}{ll} Utility(a, w_1) = 3 & Utility(b, w_1) = 3 \\ Utility(a, w_2) = 0 & Utility(b, w_2) = 4 \\ Utility(a, w_3) = 4 & Utility(b, w_3) = 0 \\ Utility(a, w_4) = 1 & Utility(b, w_4) = 1, \end{array}$$

and $V(p, w) = \text{TRUE}$ if and only if $w \in \{w_3, w_4\}$, and $V(q, w) = \text{TRUE}$ if and only if $w \in \{w_2, w_4\}$. We read p as ‘Agent a confesses’ and q as ‘Agent b confesses’.

Given $\mathfrak{M}$, it holds that

$$\begin{array}{lcl} \mathfrak{M} \models \odot^{a}_{a} p \wedge \odot^{a,b}_{a} \neg p & \text{and} & \mathfrak{M} \models \odot^{b}_{b} q \wedge \odot^{a,b}_{b} \neg q \\ \mathfrak{M} \models \odot^{a}_{a} p \wedge \odot^{a,b}_{a,b} \neg p & \text{and} & \mathfrak{M} \models \odot^{b}_{b} q \wedge \odot^{a,b}_{a,b} \neg q \\ \mathfrak{M} \not\models \odot^{a,b}_{a} (\neg p \wedge \neg q) & \text{and} & \mathfrak{M} \not\models \odot^{a,b}_{b} (\neg p \wedge \neg q). \end{array}$$

3 Two Characterizations of Moral Conflicts

Definition 7 (Characterization) Let $\mathfrak{C}$ a class of choice structures and let $\phi \in \mathfrak{L}$. Then ϕ *characterizes* $\mathfrak{C}$, if for all choice structures $\mathfrak{S}$ it holds that $\mathfrak{S} \in \mathfrak{C}$ if and only if $\mathfrak{S} \models \phi$.

3.1 Moral Conflicts of Type $\odot_{\mathcal{G}}^{\mathcal{F}_1} p \wedge \odot_{\mathcal{G}}^{\mathcal{F}_2} \neg p$

A moral conflict of type $\odot_{\mathcal{G}}^{\mathcal{F}_1} p \wedge \odot_{\mathcal{G}}^{\mathcal{F}_2} \neg p$ might occur in a choice structure $\mathfrak{S}$ if and only if there are groups $\mathcal{F}_1, \mathcal{F}_2, \mathcal{G}$ of agents in $\mathfrak{S}$ such that $\mathcal{F}_1$ is non-empty, $\mathcal{F}_2$ is non-empty, $\mathcal{F}_1$ and $\mathcal{F}_2$ are not identical, and $\mathcal{G}$ has at least two non-identical options for acting:

Theorem 1 Let $\mathfrak{C}$ be the class of choice structures $\mathfrak{S}$ such that for all $\mathcal{F}_1, \mathcal{F}_2, \mathcal{G} \subseteq A$ it holds that $\mathcal{F}_1 = \emptyset$ or $\mathcal{F}_2 = \emptyset$ or $\mathcal{F}_1 = \mathcal{F}_2$ or $Choice(\mathcal{G}) = \{W\}$. Let $p \in \mathfrak{P}$. Then

$$\bigwedge_{\mathcal{F}_1, \mathcal{F}_2, \mathcal{G} \subseteq A} \neg(\odot_{\mathcal{G}}^{\mathcal{F}_1} p \wedge \odot_{\mathcal{G}}^{\mathcal{F}_2} \neg p) \text{ characterizes } \mathfrak{C}.$$

3.2 Moral Conflicts of Type $\odot_{\mathcal{G}_1}^{\mathcal{F}} p \wedge \odot_{\mathcal{G}_2}^{\mathcal{F}} \neg p$

A moral conflict of type $\odot_{\mathcal{G}_1}^{\mathcal{F}} p \wedge \odot_{\mathcal{G}_2}^{\mathcal{F}} \neg p$ might occur in a choice structure $\mathfrak{S}$ if and only if there are groups $\mathcal{F}, \mathcal{G}_1, \mathcal{G}_2$ of agents in $\mathfrak{S}$ such that $\mathcal{F}$ is non-empty, $\mathcal{G}_1 - \mathcal{G}_2$ has at least two non-identical options for acting, $\mathcal{G}_2 - \mathcal{G}_1$ has at least two non-identical options for acting, and $\mathcal{G}_1 \cap \mathcal{G}_2$ has at least two non-identical options for acting:

Theorem 2 Let $\mathfrak{C}'$ be the class of choice structures $\mathfrak{S}$ such that for all $\mathcal{F}, \mathcal{G}_1, \mathcal{G}_2 \subseteq A$ it holds that $\mathcal{F} = \emptyset$ or $Choice(\mathcal{G}_1 - \mathcal{G}_2) = \{W\}$ or $Choice(\mathcal{G}_2 - \mathcal{G}_1) = \{W\}$ or $Choice(\mathcal{G}_1 \cap \mathcal{G}_2) = \{W\}$. Let $p \in \mathfrak{P}$. Then

$$\bigwedge_{\mathcal{F}, \mathcal{G}_1, \mathcal{G}_2 \subseteq A} \neg(\odot_{\mathcal{G}_1}^{\mathcal{F}} p \wedge \odot_{\mathcal{G}_2}^{\mathcal{F}} \neg p) \text{ characterizes } \mathfrak{C}'.$$

Part of the proof: ($\Leftarrow$) Suppose $\mathfrak{S} \notin \mathfrak{C}'$. Then there must be $\mathcal{F}, \mathcal{G}_1, \mathcal{G}_2 \subseteq A$ such that $\mathcal{F} \neq \emptyset$ and $Choice(\mathcal{G}_1 - \mathcal{G}_2) \neq \{W\}$ and $Choice(\mathcal{G}_2 - \mathcal{G}_1) \neq \{W\}$ and $Choice(\mathcal{G}_1 \cap \mathcal{G}_2) \neq \{W\}$. To prove that $\mathfrak{S} \not\models \bigwedge_{\mathcal{F}, \mathcal{G}_1, \mathcal{G}_2 \subseteq A} \neg(\odot_{\mathcal{G}_1}^{\mathcal{F}} p \wedge \odot_{\mathcal{G}_2}^{\mathcal{F}} \neg p)$, it suffices to construct a model $\mathfrak{M} = \langle \mathfrak{S}, \mathfrak{I} \rangle$ in which there is a w such that $\mathfrak{M}, w \models \odot_{\mathcal{G}_1}^{\mathcal{F}} p \wedge \odot_{\mathcal{G}_2}^{\mathcal{F}} \neg p$. We conclude from the four properties that there are at least two non-identical options K_1 and K_2 in $Choice(\mathcal{G}_1 - \mathcal{G}_2)$, at least two non-identical options L_1 and L_2 in $Choice(\mathcal{G}_2 - \mathcal{G}_1)$, and at least two non-identical options

M_1 and M_2 in $Choice(\mathcal{G}_1 \cap \mathcal{G}_2)$, and that there is an agent a in $\mathcal{F}$. Note that if $K \in Choice(\mathcal{G}_1 - \mathcal{G}_2)$ and $M \in Choice(\mathcal{G}_1 \cap \mathcal{G}_2)$, then $K \cap M \neq \emptyset$ and $K \cap M \in Choice(\mathcal{G}_1)$. Note that if $L \in Choice(\mathcal{G}_2 - \mathcal{G}_1)$ and $M \in Choice(\mathcal{G}_1 \cap \mathcal{G}_2)$, then $L \cap M \neq \emptyset$ and $L \cap M \in Choice(\mathcal{G}_2)$. We now define a suitable interpretation $\mathfrak{I} = \langle Utility, V \rangle$. First, $Utility$ is defined as follows:

$$Utility(a, w) = \begin{cases} 1, & \text{if } w \in K_1 \cap M_2 \text{ or } w \in L_2 \cap M_1 \\ 0, & \text{otherwise,} \end{cases}$$

and for all agents b in $A - \{a\}$ and for all worlds w in W, we fix $Utility(b, w) = 0$. Second, we stipulate $V(p, w) = \text{TRUE}$ if and only if $w \in K_1 \cap M_2$.

Let $\mathfrak{M} = \langle \mathfrak{S}, \mathfrak{I} \rangle$ and let $w \in W$. Now it is easy to show that (1) for all $R \in Choice(\mathcal{G}_1)$ with $R \neq K_1 \cap M_2$ it holds that $K_1 \cap M_2 \succ^{\mathcal{F}}_{\mathcal{G}_1} R$ and (2) for all $S \in Choice(\mathcal{G}_2)$ with $S \neq L_2 \cap M_1$ it holds that $L_2 \cap M_1 \succ^{\mathcal{F}}_{\mathcal{G}_2} S$. Hence, $\mathfrak{M}, w \models \odot^{\mathcal{F}}_{\mathcal{G}_1} p$ and $\mathfrak{M}, w \models \odot^{\mathcal{F}}_{\mathcal{G}_2} \neg p$. Therefore, $\mathfrak{M}, w \models \odot^{\mathcal{F}}_{\mathcal{G}_1} p \wedge \odot^{\mathcal{F}}_{\mathcal{G}_2} \neg p$. $\square$

Let us take a closer look at the countermodel to interpret it properly. The group $\mathcal{G}_1 \cap \mathcal{G}_2$ of agents cannot make a principled choice from $Choice(\mathcal{G}_1 \cap \mathcal{G}_2)$ to maximize the interest of group $\mathcal{F}$. If $\mathcal{G}_1 \cap \mathcal{G}_2$ is taken to belong to group $\mathcal{G}_1$, it has to choose option M_2 to maximize $\mathcal{F}$'s interest. On the other hand, if $\mathcal{G}_1 \cap \mathcal{G}_2$ is seen as a subgroup of group $\mathcal{G}_2$, it must rather choose option M_1 to maximize $\mathcal{F}$'s interest. Obviously, $\mathcal{G}_1 \cap \mathcal{G}_2$ cannot choose both options. The group $\mathcal{G}_1 \cap \mathcal{G}_2$ is wearing two hats here.