The Best We Can Do

Frederik Van De Putte, Allard Tamminga, & Hein Duijf

May 9, 2019

Abstract

We study a **STIT** logic for two agents i and j, augmented with three deontic constants $1, 1_i, 1_j$. The constants express, respectively, optimality for the group $\{i, j\}$, optimality for agent i, and optimality for agent j. An action X is optimal for an (individual or group) agent if and only if X is not strongly dominated by another action X' that is available to that same agent. We propose an axiomatization for this logic and we study its expressive power. In particular, we show that the deontic constants are not interdefinable, but nevertheless display significant interaction properties.

1 Introduction

Background: axiomatization of deontic STIT logic In [7], Horty develops semantics for deontic expressions of the type "agent α ought to see to it that φ" (henceforth, $O_\alpha\varphi$).[1] His analysis is cast against the background of **STIT** logic, the logic of agency that was introduced in the seminal work [3]. Roughly speaking, $O_\alpha\varphi$ is true at a state in Horty's framework if and only if every optimal action of α at that state guarantees φ. As Horty shows, the notion of "optimality" can be defined in various ways, giving rise to a range of different systems. Horty furthermore generalizes his semantics to an ought-operator for groups (henceforth, O_G) and to strategic oughts [7, Chapters 6 and 7].

Here, we focus on what Horty calls *dominance act utilitarianism* and we abstract from the temporal dimension in his models. As a result, the set of actions that are available to each agent is identified by a partition of the set of possible worlds, and optimality of an action is defined in terms of weak dominance (cf. Section 2).

Although Horty's work has received plenty of attention over the past two decades, there have been relatively few attempts to characterize his deontic logics syntactically. In [7, Chapter 6], Horty discusses a number of general inference rules, showing i.a. that the validity of those rules depends on the exact definition of optimality one is using. When it comes to dominance act utilitarianism, his main observations are negative: he shows that optimality for an individual agent α does not imply optimality for a group that contains α;

[1] An earlier version of Horty's basic ideas appeared in [6].

also the converse implication fails. Using these observations, Horty argues that certain straightforward links between individual oughts and group oughts fail, at least on his utilitarian reading of O_α, resp. O_G.

In more recent work, Tamminga and co-authors have shown that Horty's negative observations are preserved when moving to a much more narrow class of models, viz. *deontic game models*, [13] and [12]. These models are tightly linked to Schelling's *pure collaboration-games* [11] and Bacharach's *coordination contexts* [2].

In [9], a sound and complete axiomatization is given for the fragment of Horty's logic that only includes individual agency and individual ought. Murakami's axiomatization shows that for this fragment, the logic does not mirror the rich semantics Horty developed. In fact the deontic operators O_α do not display any interesting interaction principles beyond those one already has in the non-deontic fragment.[2] Murakami concludes that deontic **STIT** logic for individual agents may just as well be characterized in terms of agent-relative optimality functions, rather than using the more complex, interactive notion of optimality introduced by Horty.

Another important result was obtained by Herzig and Schwartzentruber [5], who showed that group **STIT** – without deontic operators – is not finitely axiomatizable and not decidable, as soon as one has $n \geq 3$ agents. In this case, group **STIT** contains the modal product logic $\mathbf{S5}^3$. This fact poses serious limitations to the axiomatization of deontic **STIT** logic for group agents.

This paper Our aim is to study a logic that is richer than Horty's in one respect, but more restricted in another. That is, we study a formal language with only *two* agents i and j, but with deontic *constants* 1, 1_i and 1_j. These constants express, respectively, optimality for the group $\{i, j\}$, optimality for agent i, and optimality for agent j respectively, where optimality is defined in line with Horty's deontic **STIT** logic.[3] We interpret this language using models that are equivalent to the deontic action models from [13], and hence can be seen as normal game forms equipped with a deontic optimality function for the group $\{i, j\}$. For the models we work with, Horty's O_i can be expressed in our formal language by the formula $\Box(1_i \to \varphi)$, where the operator $\Box$ quantifies over all the worlds in the model. Similarly, $O_{\{i,j\}}\varphi$ is defined as $\Box(1 \to \varphi)$.[4]

As we will show in the remainder, our language is sufficiently rich to yield a number of strong interaction principles. We thus obtain a simple, yet paradigmatic formalization of the interaction between individual and group optimality. Although Horty's negative results are preserved in this specific setting (cf.

[2]There are such principles as $O_\alpha\varphi, O_\beta\psi \vdash \Diamond(\varphi \wedge \psi)$, but these rather follow from the interaction of each single operator O_α with the corresponding agency operator $[\alpha]$ and the alethic operator $\Diamond$, together with the well-known principle of *Independence of Agents* (IOA, cf. Section 2).

[3]In fact, since we use deterministic models, optimality for the group $\{i, j\}$ coincides with deontic ideality of worlds. See Section 2 for the exact details.

[4]These definitions follow the well-known Anderson-Kangerian reduction of deontic logics to alethic modal logics. See [1] for a general introduction to this topic.

supra), we get a significant number of insightful, positive results as well.

Horty's work relies on a link between game and decision theory and the model theory of **STIT** logic. In light of this connection, the notion of optimality in our logic straightforwardly relates to what game theorists call "admissibility". Admissibility has a long tradition in decision theory (see the discussion in [8, Section 2.7]).[5] Roughly stated, optimality is the same as admissibility in games where utility functions are binary. Our logical study can be viewed as uncovering the logic of admissibility in such games. Although this is a restriction, our study reveals that this logical system is already quite complex.

Outline We will first define the formal language and semantics of our target logic. In Section 3 we propose an axiomatization for this logic and outline our main metaresults. Finally, in Section 4 we discuss the relation between optimality for an individual agent and group optimality, from the viewpoint of our results.

2 Preliminaries

Language Fix a propositional base, with atoms $\mathfrak{P} = \{p, q, \ldots\}$. We have two operators for the agents i and j: $[i]$ and $[j]$ with duals $\langle i \rangle$ and $\langle j \rangle$, and the universal box $\Box$ with dual $\Diamond$. Intuitively, $[i]$ stands for: "i sees to it that...", or "given i's current action, it is guaranteed that ...". The reading of $[j]$ is analogous. $\Box\varphi$ denotes that φ is settled true; in other words, φ is true regardless of what the agents do. We moreover have three deontic constants:

1, "the group $\{i, j\}$ performs one of its optimal actions"

1_i, "i performs one of its optimal actions"

1_j, "j performs one of its optimal actions"

The language $\mathfrak{L}$ is obtained by closing $\mathfrak{P} \cup \{1, 1_i, 1_j\}$ under the classical connectives and the aforementioned three operators. We treat $\bot$, $1, 1_i, 1_j$ $\neg$, $\vee$, $[i]$, $[j]$, and $\Box$ as primitive, the other constants, connectives and operators are defined in the usual way. For the sake of convenience, we also define three dual constants: $0 = \neg 1$, $0_i = \neg 1_i$, and $0_j = \neg 1_j$.

We now work our way towards the intended models for the logic, in three steps.

Definition 1 *An* action model *is a triple $M = \langle W, \sim_i, \sim_j, V \rangle$, where W is finite and non-empty, $\sim_i$ and $\sim_j$ are equivalence relations over W that satisfy* Independence of Agents*:*

(IOA) for all $w, w' \in W$, there is a $u \in W$ such that $w \sim_i u$ and $w' \sim_j u$

[5] Admissibility links to Savage's "sure-thing principle" [10, p. 21]; he writes: "I know of no other extralogical principle governing decisions that finds such ready acceptance."

and $V : \mathfrak{P} \cup \{1, 1_i, 1_j\} \to \wp(W)$ *is a valuation function that satisfies the condition that* $V(1) \neq \emptyset$.

Note that we assume finiteness of the models. This condition will be important later on, in order to ensure that the dominance relation over actions of the agents is smooth (i.e., there are no infinite sequences of ever better actions for a given agent). Without it, the axiomatization we provide below is not sound.

We let $|w|_i = \{v \in W \mid v \sim_i w\}$ and $\mathsf{C}_i(M) = \{|w|_i \mid w \in W\}$. $|w|_j$ and $\mathsf{C}_j(M)$ are defined analogously. Intuitively, $\mathsf{C}_i(M)$ and $\mathsf{C}_j(M)$ represent the set of actions that are available to i, resp. j in the model M. Note that the members of $\mathsf{C}_i(M)$ ($\mathsf{C}_j(M)$) are mutually exclusive and jointly exhaustive.

Condition (IOA) is well-known from the **STIT** logic literature (see [3]). It expresses the fact that actions are free, in a very strong sense: if i is able to do a given action $X \in \mathsf{C}_i(M)$, then there is no action $Y \in \mathsf{C}_j(M)$ that is incompatible with X.

The set of *cells* in M is defined as follows:

$$\mathsf{C}(M) = \{|w|_i \cap |w|_j : w \in W\}$$

Intuitively, the cells in M can be thought of as the choice of the group $\{i, j\}$ in M. If we represent an action model for two agents by a grid, the innermost squares in this grid will correspond to the cells of the model. Figure 1 provides a simple representation of an action model, where we abstract from the valuation function. In this representation, a, b, c, d, e represent worlds in the model; rows represent actions of i and columns represent actions of j.

a,b	c
d	e

Figure 1: An action model.

Definition 2 *Where* $M = \langle W, \sim_i, \sim_j, V\rangle$ *is an action model,* $\varphi, \psi \in \mathfrak{L}$*, and* $w \in W$*:*

1. *if* $\varphi \in \mathfrak{P} \cup \{1, 1_i, 1_j\}$*, then* $M, w \models \varphi$ *iff* $w \in V(\varphi)$
2. $M, w \models \neg\varphi$ *iff* $M, w \not\models \varphi$
3. $M, w \models \varphi \vee \psi$ *iff* $M, w \models \varphi$ *or* $M, w \models \psi$
4. $M, w \models \Box\varphi$ *iff for all* $v \in W$, $M, v \models \varphi$
5. $m, w \models [i]\varphi$ *iff for all* $v \in |w|_i$, $M, v \models \varphi$
6. $m, w \models [j]\varphi$ *iff for all* $v \in |w|_j$, $M, v \models \varphi$

As usual, we let $\|\varphi\|_M = \{w \in W : M, w \models \varphi\}$.

Definition 3 *A* deterministic action model *is an action model* M *where every* $X \in \mathsf{C}(M)$ *is a singleton. Equivalently,* $M = \langle W, \sim_i, \sim_j, V\rangle$ *is deterministic iff for all* $w \in W$, $|w|_i \cap |w|_j = \{w\}$.

In deterministic action models, the group $\{i, j\}$ has the ability to fully determine the world one is in. As shown in [14] and [4], deterministic action models correspond in a straightforward way to normal game forms. Moreover, as far as our language $\mathfrak{L}$ is concerned, every action model is pointwise equivalent to a deterministic action model.[6] For instance, the (non-deterministic) model represented in Figure 1 is pointwise equivalent to the deterministic model from Figure 2.

a_1	b_1	c_1	c_2
b_2	a_2	c_3	c_4
d_1	d_2	e_1	e_2
d_3	d_4	e_3	e_4

Figure 2: A deterministic action model with 16 worlds. Here, for each $x \in \{a, b, c, d, e\}$, worlds x_i correspond to world x in Figure 1.

In a deterministic action model M, one can interpret the set $\|1\|_M$ as the set of best worlds *simpliciter*, since every action of $\{i, j\}$ corresponds to a unique world. It should however be kept in mind that when we speak of a "best" world, this is a world in which the group $\{i, j\}$ is doing one of its best actions.

So far, we allow for models where the interpretation of 1_i and 1_j can be any set of worlds. We will now define the class of deterministic action models in which the interpretation of 1_i (1_j) coincides with the interpretation of optimality as "not being strongly dominated by a different action of i (j), with respect to the best worlds". To spell this out for deterministic models, we need some more notation.

Let $X, Y \in \mathsf{C}_i(M)$. Then X weakly dominates Y, in symbols: $X \sqsubseteq_i Y$, iff for all $Z \in \mathsf{C}_j(M)$: if $Y \cap Z \subseteq \|1\|_M$, then $X \cap Z \subseteq \|1\|_M$. Likewise, where $X, Y \in \mathsf{C}_j(M)$, $X \sqsubseteq_j Y$ iff for all $Z \in \mathsf{C}_i(M)$: if $Y \cap Z \subseteq \|1\|_M$, then $X \cap Z \subseteq \|1\|_M$. Strong dominance for $\alpha \in \{i, j\}$, denoted by $\sqsubset_\alpha$, is defined from $\sqsubseteq_\alpha$ in the standard way. Where $\alpha \in \{i, j\}$, we say that $X \in \mathsf{C}_\alpha(M)$ is *optimal for α in M* iff there is no $X' \in \mathsf{C}_\alpha(M)$ such that $X' \sqsubset X$.

Remark that our terminology here follows that of Horty [7], and is different from standard terminology in game theory. The game-theoretic concept of "weak dominance" coincides with what we call strong dominance, and weak domination is simply absent.

Definition 4 *M is a* deontic action model *iff M is a deterministic action model and satisfies the following two conditions on V:*

(D_i) $w \in V(1_i)$ *iff* $|w|_i$ *is optimal for i in M*
(D_j) $w \in V(1_j)$ *iff* $|w|_j$ *is optimal for j in M*

A deontic action model is hence a deterministic model in which 1_i (1_j) is true at a point w in the model if and only if i (j) does one of its optimal actions

[6] See [16]. This result is generalized to (finite and infinite) models for n agents in [17].

at w. Figures 3 and 4 represent two deterministic action models, this time with an explicit representation of the valuation function for the deontic constants $1, 1_i, 1_j$. Here, we put a given constant on top of a column (resp., before a row) to indicate that this constant is true in all worlds of that column (row). The first of these two models violates condition (D_j) from Definition 4 and hence is not a deontic action model. The second is a proper deontic action model.

	1_j	0_j	0_j
0_i	0	0	0
1_i	1	0	0
1_i	0	1	0

Figure 3: A deterministic action model that is not a deontic action model.

	1_j	1_j	0_j
0_i	0	0	0
1_i	1	0	0
1_i	0	1	0

Figure 4: A deontic action model.

As usual, φ is *valid* iff for every deontic action model M and every w in the domain of M, it holds that $M, w \models \varphi$.

3 Axiomatization and other meta-results

Table 1 provides a sound axiomatization for our logic. The first part of this table contains axioms that are well-known from the general study of **STIT** logic. The second part covers the logical behavior of the deontic constants. We further comment on this second part below.

In the remainder we use (BA), (BND), and (NBD) to refer to either of the two indexed versions of the respective axioms. (BA) expresses that 1_i, resp. 1_j is either true of all worlds that belong to some available action of i (j), or of no world that belongs to that action. (PB) expresses the condition that $V(1) \neq \emptyset$. Importantly, (PB) is independent of the other axioms (one can have all the others without PB, just by allowing for models with no permissible world).

(BND) is implied by the fact that if some action is best for a given agent, then it is not dominated by any other action of that same agent. For instance, if the negation of the right hand side of (BND_i) holds at a world w, then this implies that $|w|_i$ is strongly dominated by some other available action of i in the model at hand.

(NBD_i) and (NBD_j) are a shorthand for infinitely many (distinct) axiom schemata — here, n is any natural number larger than 0. These axioms follow

Non-deontic part:

(MP)	Modus ponens
(CL)	propositional classical logic
(NEC)	Necessitation for $[i], [j], \Box$
(S5)	**S5** for $[i], [j], \Box$
(GM)	$\Box\varphi \to [i]\varphi$
(IOA)	$(\Diamond[i]\varphi \wedge \Diamond[j]\psi) \to \Diamond([i]\varphi \wedge [j]\psi)$

Deontic part:

(BA_i)	$1_i \to [i]1_i$
(BA_j)	$1_j \to [j]1_j$
(PB)	$\Diamond 1$
(BND_i)	$1_i \to [i]((0 \wedge \langle j\rangle(1 \wedge \varphi)) \to \langle i\rangle(1 \wedge \langle j\rangle(0 \wedge \langle i\rangle\varphi)))$
(BND_j)	$1_j \to [j]((0 \wedge \langle i\rangle(1 \wedge \varphi)) \to \langle j\rangle(1 \wedge \langle i\rangle(0 \wedge \langle j\rangle\varphi)))$
(NBD_i)	$(0_i \wedge [i]([j]\varphi \to 1) \wedge \langle i\rangle([j]\psi_1 \wedge 1) \wedge \ldots \wedge \langle i\rangle([j]\psi_n \wedge 1)) \to$ $\Diamond(1_i \wedge [i]([j]\varphi \to 1) \wedge \neg[j]\varphi \wedge 1 \wedge \langle i\rangle([j]\psi_1 \wedge 1) \wedge \ldots \wedge \langle i\rangle([j]\psi_n \wedge 1))$
(NBD_j)	$(0_j \wedge [j]([i]\varphi \to 1) \wedge \langle j\rangle([i]\psi_1 \wedge 1) \wedge \ldots \wedge \langle j\rangle([i]\psi_n \wedge 1)) \to$ $\Diamond(1_j \wedge [j]([i]\varphi \to 1) \wedge \neg[i]\varphi \wedge 1 \wedge \langle j\rangle([i]\psi_1 \wedge 1) \wedge \ldots \wedge \langle j\rangle([i]\psi_n \wedge 1))$

Table 1: Axiom schemas for the logic of deontic action models.

from the fact that if the current action X of a given agent is not best, then it must be dominated by some other action Y of the same agent, where Y is in fact optimal for i.

Theoremhood ($\vdash \varphi$) is defined in the usual way, i.e. as derivability from all instances of the above axiom schemata, using modus ponens and necessitation. In the remainder, we use $\mathbf{L}$ to refer to the resulting logic, and where $m \in \mathbb{N}$, we use $\mathbf{L}_m$ to refer to the logic obtained by having the axiom schemata (NBD_i) and (NBD_j) for all $n \leq m$.

Our main metaresults are summarized by the following two theorems:

Theorem 1 (Soundness) *If* $\vdash \varphi$*, then* $\models \varphi$*.*

Theorem 2 (Finite fragments) *For every* $m \in \mathbb{N}$*:* $\mathbf{L}_m \subset \mathbf{L}$*.*

We moreover conjecture:[7]

Conjecture 1 (Completeness) *If* $\models \varphi$*, then* $\vdash \varphi$*.*

[7]This conjecture is suggested by the fact that we can derive particularly strong theorems using this axiomatization, cf. Section 4. However, notwithstanding serious efforts, we have failed to complete the proof of both conjectures so far.

4 Individual and Group Optimality

Let us now return to the issue that motivated this research, i.e. the relation between individual and group optimality and the possibility of characterizing this relation at the syntactic level. We start with an important result concerning the expressive power of our formal language.

Theorem 3 (Non-definability) *Each of the following hold:*

1. 1 *is not definable in the fragment of* $\mathfrak{L}$ *without* 1.
2. 1_i *is not definable in the fragment of* $\mathfrak{L}$ *without* 1_i.
3. 1_j *is not definable in the fragment of* $\mathfrak{L}$ *without* 1_j.

Each item of Theorem 3 can be proven by a bisimulation argument, drawing on examples from [13, 12]. The theorem can be interpreted as saying that logically speaking, claims concerning collective obligations cannot be reduced to claims concerning individual obligations — at least if one sticks to this specific interpretation of optimality, and the notion of obligation that correlates with it. This is hence a significant strengthening of Horty's earlier result that shows that individual oughts do not imply collective oughts and vice versa (cf. Section 1).

The fact that the deontic constants $1, 1_i, 1_j$ are not interdefinable does however not mean that they display no interaction. In contrast to Murakami's axiomatization of deontic **STIT** logic [9], the logic displays strong interaction principles. We will now go over a few of the derivable theorems of the logic in order to illustrate this fact.[8]

Observation 1 *Each of the following hold:*

1. $\vdash [i](\langle j\rangle 1 \to 1) \to 1_i$
2. $\vdash \Diamond[i](\langle j\rangle 1 \to 1) \to (1_i \to [i](\langle j\rangle 1 \to 1))$

Observation 1.1 can be derived from (NBD_i), by putting $\varphi = \langle j\rangle 1$, and relying on **S5**-properties of $[j]$. The antecedent of this theorem states that, there is an action X available to agent i such that, for every action Y available to j, if Y contains a 1-world, then the world in $X \cap Y$ is also a 1-world. It can be easily verified that if this is the case, then X weakly dominated every other action of i; hence X is optimal for i.

Observation 1.2 can be derived using (BND_i). That is, assume for contradiction that 1_i, $\Diamond[i](\langle j\rangle 1 \to 1)$, and $\langle i\rangle(\langle j\rangle 1 \wedge 0)$. From there, we reason as follows:

1.	$\langle i\rangle(\langle j\rangle 1 \wedge \Diamond[i](\langle j\rangle 1 \to 1) \wedge 0)$	by S5-properties of $\Box$ and (GM)
2.	$\langle i\rangle(\langle j\rangle 1 \wedge \langle j\rangle\langle i\rangle[i](\langle j\rangle 1 \to 1) \wedge 0)$	since[9] $\vdash \Diamond\varphi \leftrightarrow \langle j\rangle\langle i\rangle\varphi$

[8] We rely freely on normal modal logic properties in the derivations of these theorems.

[9] This theorem, also know as the "Church-Rosser Property" is known to hold for **STIT** logic [16]. One half of the equivalence is easy, using (GM). For the other half, assume $\Diamond\varphi$ and $[i][j]\neg\varphi$. By the first assumption we have $\Diamond\langle j\rangle\varphi$. Hence, $\Diamond[j]\langle j\rangle\varphi$. From the second assumption we get $\Diamond[i][j]\neg\varphi$. If we now apply (IOA), we get a contradiction.

3. $\langle i\rangle([j]\langle j\rangle 1 \wedge \langle j\rangle[i](\langle j\rangle 1 \to 1) \wedge 0)$ by S5-properties of $[j]$, resp. $[i]$
4. $\langle i\rangle(\langle j\rangle(\langle j\rangle 1 \wedge [i](\langle j\rangle 1 \to 1)) \wedge 0)$ since $[j]\psi, \langle j\rangle\tau \vdash \langle j\rangle(\tau \wedge \psi)$
5. $\langle i\rangle(\langle j\rangle(\langle j\rangle 1 \wedge (\langle j\rangle 1 \to 1) \wedge [i](\langle j\rangle 1 \to 1)) \wedge 0)$ T-schema for $[i]$, (RE)
6. $\langle i\rangle(\langle j\rangle(1 \wedge [i](\langle j\rangle 1 \to 1)) \wedge 0)$ modus ponens in the scope of $\langle j\rangle$

Then, putting $\varphi = [i](\langle j\rangle 1 \to 1)$, apply ($\mathrm{BND}_i$) to derive:

$$\langle i\rangle(1 \wedge \langle j\rangle(0 \wedge \langle i\rangle(\langle j\rangle 1 \to 1))$$

This then allows you to derive the contradiction $\langle i\rangle(1 \wedge 0)$ in a few steps.

Together, items 1 and 2 of Observation 1 entail a necessary and sufficient condition for i-optimality, under the specific circumstance where i can make a choice that weakly dominates every other choice:

Observation 2 $\vdash \Diamond[i](\langle j\rangle 1 \to 1) \to \Box(1_i \leftrightarrow [i](\langle j\rangle 1 \to 1))$

In a nutshell, when there is an action of an agent α that weakly dominates every other action α, optimality for that agent *is* definable in terms of optimality for $\{i, j\}$ and the non-deontic language of **STIT** logic. This means that our non-definability results (cf. Theorem 3) are only possible in view of the presence of genuine coordination problems.

Observation 3 $\vdash \Diamond[i](\langle j\rangle 1 \to 1) \to \Box((1_i \wedge 1_j) \to 1)$

Observation 3 gives us a sufficient (but non-necessary) condition for when optimality carries over from the members i, j to the group $\{i, j\}$. Under the assumption that one of the two agents has an individual action available that weakly dominates any other individual action available to her, if each individual agent performs an optimal individual action, then the group performs an optimal group action. Or, equivalently, under this assumption, if the group does not perform an optimal group action, then at least one of its members does not perform an optimal individual action. The theorem can be derived from Observation 1.2 and the fact that $\vdash 1_j \to \langle j\rangle 1$, itself a consequence of (BND_j).

Observation 4 *Each of the following hold:*

1. $\vdash \Diamond 1_i$
2. $\vdash 1 \to \langle j\rangle(1_i \wedge 1)$
3. $\vdash \Diamond(1 \wedge 1_i \wedge 1_j)$

Observation 4.1 can be derived from (NBD_i): if the current action is not optimal for i, then (NBD_i) entails that there is some other optimal action for i that strictly dominates it. It can be seen as a variant of the Kantian "ought implies can" principle: each agent α can always fulfill her obligation, in the sense that it can perform an α-optimal action. Note that from this theorem and its counterpart for j, using (IOA), we can derive $\Diamond(1_i \wedge 1_j)$, expressing that it is possible that all agents fullfill their obligations.

Observation 4.2 follows from (NBD_i). Finally, Observation 4.3 follows from Observation 4.2, (PB), and (BA_i):

1. $\Diamond 1$ (PB)
2. $\Diamond\langle j\rangle(1 \wedge 1_i)$ by Observation 4.2
3. $\Diamond\langle j\rangle(\langle i\rangle(1 \wedge 1_j) \wedge 1_i)$ by the counterpart of Observation 4.2 for j
4. $\Diamond\langle j\rangle(\langle i\rangle(1 \wedge 1_j) \wedge [i]1_i)$ by (BA_i)
5. $\Diamond\langle j\rangle\langle i\rangle(1 \wedge 1_j \wedge 1_i)$ NML property: $\langle i\rangle\varphi \wedge [i]\psi \vdash \langle i\rangle(\varphi \wedge \psi)$
6. $\Diamond\Diamond\Diamond(1 \wedge 1_j \wedge 1_i)$ converse of (GM) for $\langle i\rangle$ and for $\langle j\rangle$
7. $\Diamond(1 \wedge 1_j \wedge 1_i)$ by the 4-axiom for $\Box$

Observation 4.3 expresses the fact that, in a simple coordination game, there are always optimal individual actions available for agents i and j such that the combination of those actions is an optimal group action. Consequently, *if* the agents could settle on a fixed plan for coordination, they will always be able to solve the coordination problem. This however does not go against the earlier mentioned result that one can never reduce optimality for the individuals to optimality for the group, or conversely. The fact that there *is* a solution to the coordination game by no means implies that the agents can coordinate in order to carry out that solution. After all, communication could be impossible and agreement problematic.

References

[1] Lennart Åqvist. *Deontic Logic*, volume 8 of *Handbook of Philosophical Logic*, chapter 4, pages 147–264. Kluwer Academic Publishers, 2 edition, 2002.

[2] M. Bacharach. *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton University Press, Princeton: NJ, 2006.

[3] N. Belnap, Perloff M., Xu M., and Bartha P. *Facing the Future: Agents and Choice in Our Indeterminist World*. Oxford University Press, 2001.

[4] Roberto Ciuni and John Horty. *Stit Logics, Games, Knowledge, and Freedom*, volume 5 of *Outstanding Contributions to Logic*, chapter 23, pages 631–656. Springer International Publishing, August 2014.

[5] Andreas Herzig and François Schwarzentruber. Properties of logics of individual and group agency. In Carlos Areces and Robert Gobldblatt, editors, *Advances in Modal Logic*. College Publications, 2008.

[6] John F. Horty. *Deontic Logic, Agency and Normative Systems: ΔEON '96: Third International Workshop on Deontic Logic in Computer Science, Sesimbra, Portugal, 11 – 13 January 1996*, chapter Combining Agency and Obligation (Preliminary Version), pages 98–122. Springer London, London, 1996.

[7] John F. Horty. *Agency and Deontic Logic*. Oxford University Press, New York, 2001.

[8] Elon Kohlberg and Jean-Francois Mertens. On the strategic stability of equilibria. *Econometrica*, 54(4):1003–37, 1986.

[9] Yuko Murakami. Utilitarian deontic logic. In *Advances in Modal Logic*, volume 5, pages 211–230, 2005.

[10] L.J. Savage. *The Foundations of Statistics*. Wiley & Sons, 1954.

[11] Thomas Schelling. *The Strategy of Conflict*. Harvard University Press, Cambridge, Massachusetts, 1960.

[12] Allard Tamminga and Hein Duijf. Collective obligations, group plans and individual actions. *Economics and Philosophy*, 33(2):187–214, 2017.

[13] Allard Tamminga and Frank Hindriks. The irreducibility of collective obligations. Unpublished Manuscript.

[14] Paolo Turrini. Agreements as norms. In Thomas Ågotnes, Jan M. Broersen, and Dag Elgesem, editors, *Deontic Logic in Computer Science - 11th International Conference*, pages 31–45. Springer, 2012.

[15] Alasdair Urquhart. Decidability and the finite model property. *Journal of Philosophical Logic*, 10(3):367–370, Aug 1981.

[16] Johan van Benthem and Eric Pacuit. *Connecting Logic of Choice and Change*, volume 2 of *Outstanding Contributions to Logic*, chapter 14, pages 291–314. Springer International Publishing, 2014.

[17] Frederik Van De Putte, Allard Tamminga, and Hein Duijf. Doing without nature. In *LORI VI*, 2017.