A Deontic Logic for Socially Optimal Norms

Jan Broersen, Rosja Mastop, John-Jules Ch. Meyer, Paolo Turrini

Universiteit Utrecht
Email: paolo@cs.uu.nl

Abstract

The paper [a] discusses the interaction properties between preference and choice of coalitions in a strategic interaction. A language is presented to talk about the conflict between coalitionally optimal and socially optimal choices. Norms are seen as social constructions that enable to enforce socially desirable outcomes.

[a]This paper has been presented at the Ninth International Workshop on Deontic Logic in Computer Science (DEON'08) and it is to be found in the conference proceedings (see [4]).

1 Introduction

One fundamental issue of social choice theory [1] is how to aggregate the preferences of individual agents in order to form decisions to be taken by society as a whole. However, once we want to take into account the capabilities of agents, as we do in Multi Agent Systems, mere social choice functions are not enough to explain how and (especially) why individual interests are aggregated in the way they are. In this context, norms should be seen as social constructions that enable us to enforce socially desirable outcomes [5].

In particular there are situations in which individual preferences are not compatible and coalitions compete to achieve a given social order. A typical case is that of an agent's capability to positively or negatively affect the realization of other agents' preferences. In our paper we will view the enactment of norms as aimed at the regulation of such interactions. By enacting a norm we mean *the introduction of a normative constraint on individual and collective choices in a Multi Agent System.*

We are specifically concerned with cases where the collective perspective is at odds with the individual perspective. That is, cases where we think that letting everybody pick their own best action regardless of others' interest gives a non-optimal result. The main question we are dealing with is then: how do we determine which norms, if any, are to be imposed?

To answer this question, the paper presents a language to talk about the conflict between coalitionally optimal and socially optimal choices, and it expresses deontic notions referring to such circumstances.

1.0.1 Motivating Example

Let us consider a situation (Table 1) in which two players have the possibility of passing believed (truth) or disbelieved information (lie). If both players do not lie, they share their information, being both better off. If they both lie, they do not receive any advantage. But the worst case for a player is the one in which he does not lie and the opponent does.

In this situation, a legislator that wants to achieve the socially optimal state (players do not lie), should declare that lying is forbidden, thereby labeling the combinations of moves (lie, lie), (lie, truth) and (truth, lie) as violations.

The lying matrix is nothing but a Prisoner Dilemma [11], that is an interactive situation in which the advantages of cooperation are overruled by the incentive for individual players to defect. In Prisoner Dilemmas, individually rational players have no incentive to cooperate, because defecting is better for a player considering all possible answers of the opponent. Note that cooperation is in the interest of the players themselves, since they would be better off than if they had pursued the unique Nash Equilibrium [11], ending up in the (lie, lie) state. However it is by no means clear that players should not pursue their own interest. In fact once we reach a state in which one player lies and the other does not lie, we cannot move to any other state without one of the two players being worse off.

1.0.2 A deontic logic for efficient interactions

Once we view a deontic language as regulating a Multi Agent System, we can say that a set of commands promote a certain interaction, prohibiting certain others. Following this line of reasoning it is possible, given a notion of optimality or efficiency, to provide a set of deontic formulas that agree with such notion, as we have done with Pareto Optimality.

What we do then is to provide a deontic language for all possible interactions, based on an underlying notion of optimality. This is quite a difference from the legal codes that we can find in a certain society, where norms are either explicitly and specifically formulated and written down in law books, contracts, etc., or are left implicit in the form of promises, values or mores [5]. The obligations and prohibitions in our system result from one general norm saying that all actions of sub-groups that do not take into account the interests of the society as a whole, are forbidden. Then, one way to use our logic is to derive obligations, permissions and prohibitions from conflicting group preferences, and use these as *suggestions* for norm introduction in the society.

We do not claim that the meaning of these operators, as studied in deontic logic, corresponds to our semantics, but rather we claim that when people make new norms they should choose those norms on the basis of the economical order behind them [17].

In order to represent abilities of agents we employ coalition logic [13], and we model an agent's preferences as a preorder on the domain of discourse. To model optimal social norms we introduce a generalization of the economical notion of Strong Pareto Efficiency (see for instance [11]), described as those sets of outcomes from which the grand coalition (i.e. the set of all agents taken together) has no interest to deviate. Our generalization consists of the fact that we do not make the assumption that these outcomes are singletons. In particular (unless specified) we do not make the assumption of playability described in [13], according to which the set of all agents can bring about any realizable outcome of the system. We consider then the elements of the complement of the efficient choices, i.e. all those that are not optimal, and we build the notion of obligation, prohibition and permission on top of them.

We postpone to future work all considerations about the effectivity of the norm, that is, all considerations about how, to what extent and in what way, the norm influences the behaviour of the agents involved.

As system designers, our aim is then to construct efficient social procedures that can guarantee a socially desirable property to be reached. We think that normative system design is at last a proper part of the Social Software enterprise [12].

Row \ Column	Truth	Lie
Truth	$(3, 3)$	$(0, 4)$
Lie	$(4, 0)$	$(1, 1)$

Table 1: Lying or not lying

The paper is structured as follows: In the first section we introduce the notions of effectivity and preference, discuss its relevant properties with respect to the problem of finding optimal social norms, and introduce the notion of domination, Pareto Efficiency and violation. In the second part we describe the syntax, the structures and the interpretation of our language. In the third part we discuss the deontic and collective ability modalities and their properties, and compare them with classical deontic and agency logics; moreover we discuss the introduction of further constraints in the models, in particular playability of the effectivity function. We show some examples to give the flavour of the situations we are able to capture with our formalism. A discussion of future work will follow and a summary of the present achievements will conclude the paper.

2 Effectivity and preference

We start by defining some concepts underlying the deontic logic of this paper. They concern the *power* and the *preferences* of collectives. We begin with the first of these, by introducing the concept of a dynamic effectivity function, adopted from [13].

2.1 Effectivity

Definition 1 (Dynamic Effectivity Function)
Given a finite set of agents Agt and a set of states W, a dynamic effectivity function *is a function $E : W \to (2^{Agt} \to 2^{2^W})$.*

Any subset of Agt will henceforth be called a *coalition*. For elements of W we use variables $u, v, w, \ldots$; for subsets of W we use variables $X, Y, Z, \ldots$; and for sets of subsets of W (i.e., elements of 2^{2^W}) we use variables $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \ldots$. The elements of W are called 'states' or 'worlds'; the subsets of Agt are called 'coalitions'; the sets of states $X \in E(w)(C)$ are called the 'choices' of coalition C in state w. The set $E(w)(C)$ is called the 'choice set' of C in w. The complement of a set $\overline{X}$ or of a choice set $\overline{\mathcal{X}}$ are calculated from the obvious domains.

A dynamic effectivity function assigns, in each world, to every coalition a set of sets of states. Intuitively, if $X \in E(w)(C)$ the coalition is said to be able to *force* or *determine* that the next state after w will be some member of the set X. If the coalition has this power, it can thus prevent that any state *not* in X will be the next state, but it might not be able to determine *which* state in X will be the next state. Possibly, some other coalition will have the power to refine the choice of C.

Many properties can be attributed to dynamic effectivity functions. An extensive discussion of them can be found in [13]. For what follows we do not need all the properties that may be considered reasonable for effectivity. However the following properties seem to be minimally required:

1. coalition monotonicity: for all X, w, C, D, if $X \in E(w)(C)$ and $C \subseteq D$, then $X \in E(w)(D)$;
2. regularity: for all X, w, C, if $X \in E(w)(C)$, then $\overline{X} \notin E(w)(\overline{C})$;
3. outcome monotonicity: for all X, Y, w, C, if $X \in E(w)(C)$ and $X \subseteq Y$, then $Y \in E(w)(C)$;
4. inability of the empty coalition: for all w, $E(w)(\emptyset) = \{W\}$

If a dynamic effectivity function has these properties, it will be called *coherent*.

The first property says that the ability of a coalition is preserved by enlarging the coalition. In this sense we do not allow new members to interfere with the preexistent capacities of a group of agents. The second property says that if a coalition is able to force the outcome of an interaction to belong to a particular set, then no possible combinations of moves by the other agents can prevent this to happen. We think that regularity is a key property to understand the meaning of ability. If an agent is properly able to do something this means that others have no means to prevent it. The third property says that if a coalition is able to force the outcome of the interaction to belong to a particular set, then that coalition is also able to force the outcome to belong to all his supersets. Outcome monotonicity is a property of all effectivity functions in coalition logic, which is therefore a monotonic modal logic [13]. The last condition is the "Inability of the Empty Coalition". As notice also by [2] such properties forces the coalition modality for the empty coalition to be universal: intuitively the empty coalition cannot bring about non-trivial consequences.

Proposition 1 *If the effectivity function is coherent then all coalitional effectivity functions are nonempty and do not contain the empty set.*

The last property ensures that the choice of the empty coalition is always the largest possible one. This property imposes that for all $C, w, E(w)(C) \neq \emptyset$ (by coalition monotonicity) and that $\emptyset \notin E(w)(C)$ (by regularity).

2.2 Preference

Once we have defined the notion of effectivity, we also need to make reference to the preferences of coalitions. The notion of preference in strategic interaction can be understood and modeled in many ways [16]. However we believe that in strategic reasoning players need to have preferences over the possible outcomes of the game. Thus those are the preferences that constitute our main concern. Nevertheless we know from the properties of effectivity as described above, that coalitions may have different abilities at different states, in particular the grand coalition of agents may gain or lose power while changing a state: the effectivity is actually a dynamic effectivity.

The claim is thus that agents do have a fixed ordering over the domain of discourse (what we call *preferences*), and that they generate their strategic preference considering where the game may end (called *domination*). We are going to define both, discussing their properties.

We start from a preference relation for individuals over states working our way up to preferences for coalitions over sets. To do so, we start from an order on singletons, and we provide some properties to lift the relation to sets.

Definition 2 (Individual preferences for states) *A preference ordering $(\geq_i)_{i \in Agt}$ consists of a partial order (reflexive, transitive, antisymmentric) $\geq_i \subseteq W \times W$ for all agents $i \in Agt$, where $v \geq_i w$ has the intuitive reading that v is 'at least as nice' as w for agent i. The corresponding strict order is defined as usual: $v >_i w$ if, and only if, $v \geq_i w$ and not $w \geq_i v$.*

Definition 3 (Individual preferences for sets of states) *Given a preference ordering $(\geq_i)_{i \in Agt}$, we lift it to an ordering on nonempty sets of states by means of the following principles.*

1. *$\{v\} \geq_i \{w\}$ iff $v \geq_i w$;* *(Singletons)*
2. *$(X \cup Y) \geq_i Z$ iff $X \geq_i Z$ and $Y \geq_i Z$;* *(Left weakening)*
3. *$X \geq_i (Y \cup Z)$ iff $X \geq_i Y$ and $X \geq_i Z$.* *(Right weakening)*

We do not give a comprehensive specification of the logical properties of preference relations for coalitions, because this would not be relevant for the remainder of this paper. Different types of interaction may warrant the assumption of different properties for such a relation. Nevertheless, these are some properties that seem minimally required for calling some relation a preference relation. The first ensures that preferences are copied to possible choices. The properties of left and right weakening ensure a lifting from singletons to sets.

The lifting enables us to deal with preference under uncertainty or indeterminacy. The idea is that if an agent were ever confronted with two choices X, Y he would choose X over Y provided $X >_i Y$ [1].

Preferences do not consider any realizability condition, they are simply basic aspirations of individual agents, on which to construct a more realistic order on the possible outcomes of the game, which are by definition dependent on what all the agents can do together.

Out of agents' preferences, we can already define a classical notion of Pareto Efficiency.

Definition 4 (Strong Pareto efficiency) *Given a choice set $\mathcal{X}$, a choice $X \in \mathcal{X}$ is* Strongly Pareto efficient *for coalition C if, and only if, for no $Y \in \mathcal{X}$, $Y \geq_i X$ for all $i \in C$, and $Y >_i X$ for some. When $C = Agt$ we speak of* Strong Pareto Optimality*.*

We will use the characterization of Pareto Efficiency and Optimality to refer to the notions we have just defined, even though the classical definitions (compare with [11]) are weaker.

The last definition is clearer when we consider the case $\mathcal{X} = E(w)(C)$, but it is formulated in a more abstract way in order to smoothen the next two definitions.

Proposition 2 *Given the preference relation over choices $\geq_i$, and taking A, B in a choice set $\mathcal{X}$ of a coaltion C, with $PE(A)$ to indicate that the choice A is Strongly Pareto Efficient in $\mathcal{X}$, Strong Pareto Efficiency is monotonic, that is $A \subseteq B$ implies that $PE(B)$ whenever $PE(A)$.*

Proof Suppose $A \subseteq B$ and $PE(A)$ and suppose it is not the case that $PE(B)$. This means that there is a choice X in the choice set $\mathcal{X}$ of C such that $X >_i B$ for some $i \in C$. But being $A \subseteq B$ this would imply that $X >_i A$, contradicting the assumption that $PE(A)$.

Pareto Efficiency is usually defined disregarding the strategies of the players. Nevertheless, once we claim that the outcome of an interaction need not be a singleton, we need to adapt our evaluation of efficiency to such an assumption.

We now construct a preference relation on choices. To do so we first need to look at the interaction that agents' choices have with one another.

Definition 5 (Subchoice) *If E is an effectivity function, and $X \in E(w)(\overline{C})$, then the* X-subchoice set *for C in w is given by $E^X(w)(C) = \{X \cap Y \mid Y \in E(w)(C)\}$.*

As an example, let us take Table 1. Consider expressions of the form (Lie_C) to be intended as the set of worlds that make the proposition Lie_C true, with the obvious reading. In our example we have for instance the following cases:

- $E^{(Lie_C)}(w)(R) = \{(Lie_C \wedge Lie_R), (Lie_C \wedge Truth_R)\}$
- $E^{(Truth_C)}(w)(R) = \{(Truth_C \wedge Lie_R), (Truth_C \wedge Truth_R)\}$

[1] Preference lifting in interaction is also addessed by Gardenfors in [6]. A modal account of it is given in Fenrong Liu's PhD thesis [9].

Subchoices allow us to reason on a restriction of the game and to consider possible moves looking from a coalitional point of view, i.e. what is best for a coalition to do provided the others have already moved.

When agents interact therefore they make choices on the grounds of their own preferences. Nevertheless the moves at their disposal need not be all those that the grand coalition has. We can reasonably assume that preferences are filtered through a given coalitional effectivity function. That is we are going to consider what agents prefer among the things they can do.

Definition 6 (Domination) *Given an effectivity function E, X is* undominated *for C in w (abbr. $X \rhd_{C,w}$) if, and only if, (i) $X \in E(w)(C)$ and (ii) for all $Y \in E(w)(\overline{C})$, $(X \cap Y)$ is Pareto efficient in $E^Y(w)(C)$ for C.*

The idea behind the notion of domination is that if X' and X'' are both members of $E(w)(C)$ then, in principle, C will not choose X'', if X' dominates X''. This property ensures that a preference takes into account the possible moves of the other players. This resembles the notion of Individual Rationality in Nash solutions [11], according to which an action is chosen reasoning on the possible moves of the others.

Continuing our example, we have the following cases:

- $(Lie_R) \rhd_{R,w}$ for any w.
- $(Lie_C) \rhd_{C,w}$ for any w.
- not $(Lie_C, Lie_R) \rhd_{Agt,w}$

The preceding three definitions capture the idea that 'inwardly' coalitions reason Pareto-like, and 'outwardly' coalitions reason strategically, in terms of strict domination. A coalition will choose its best option given all possible moves of the opponents. Looking at the definition of Optimality we gave, we can see that undomination collapses to individual rationality when we only consider individual agents, and to Pareto efficiency when we consider the grand coalition of agents.

Proposition 3

$X \rhd_{Agt,w}$ iff X is a standard Pareto Optimal Choice in w.

$X \rhd_{i,w}$ iff X is a standard Dominating Choice in w for i.

Proof For the first, notice that since $E(w)(\emptyset) = \{W\}$ (i.e., the empty coalition has no powers), then X is undominated for Agt in w iff it is Pareto efficient in $E(w)(Agt)$ for Agt (i.e., it is Pareto optimal in w). The second is due to the restriction of undomination to singleton agents.

Nevertheless, in our framework, domination is a relation between the choice sets of a given coalition. This approach looks different from the standard one (see for instance Osborne and Rubinstein [11]) that considers instead domination as a property of states.

Proposition 4 *Game-theoretical domination is expressible in our framework.*

It is possible to rewrite a domination of a state x over a state y as the domination of the choice $\{x\}$ over $\{y\}$, making it a particular case of our definition.

Proposition 5 *Undomination is monotonic, that is for X in $E(C)(w)$ for some C, w, if $X \subseteq Y$ and $X \rhd_{C,w}$ then $Y \rhd_{C,w}$.*

which follows from monotonicity of Pareto Optimality for choices and outcome monotonicity.

2.2.1 Violation

The fundamental idea of this work is that an efficient way to impose normative constraints in a Multi Agent System is to look at the optimality of the strategic interaction of such system. In particular the presence of possible outcomes in which agents could not unanimously improve (Pareto Efficient) can be a useful guide line for designing a new set of norms to be imposed.

Following this line we define a set of violation sets as the set of those choices that are not a Pareto Efficient interaction.

Definition 7 (Violation) *If E is an effectivity function and $C \subseteq C'$, then the choice $X \in E(w)(C)$ is a violation by C towards C' in w ($X \in VIOL_{C,C',w}$) iff there is a $Y \in E(w)(C' \setminus C)$, s.t. $(X \cap Y)$ is not undominated for C' in w.*

In words, X is a violation if it is not safe for the other agents, in the sense that not all the moves at their disposal yield an efficient outcome.

We indicate with $VIOL_{C,w}$ the set violations by C at w towards Agt[2].

Proposition 6 *If C=C' then a violation is a dominated choice; If C=C'=Agt a violation is a Pareto inefficient choice.*

If we consider the Prisoner Dilemma of Table 1 the following holds:

- $(Lie_R) = VIOL_{R,w}$ for any w, since $(Lie_R \wedge Lie_C)$ is not Agt- undominated;
- $(Lie_C \wedge Lie_R) = VIOL_{Agt,w}$ for any w, since not Pareto Efficient.

We can observe here that any choice made by a single agent is a violation. The reason why it is so has to be found in the form of the game, that requires the grand coalition to form for an efficient outcome to be forced.

3 Logic

We now introduce the syntax of our logic, an extension of the language of coalition logic [13] with modalities for permission, prohibition and obligation, and a modality for rational choice.

3.1 Language

Let Agt be a finite set of agents and $Prop$ a countable set of atomic formulas. The syntax of our logic is defined as follows:

$$\phi ::= p|\neg\phi|\phi \vee \phi|[C]\phi|P(C,\phi)|F(C,\phi)|O(C,\phi)|[rational_C]\phi$$

where p ranges over $Prop$ and C ranges over the subsets of Agt. The other boolean connectives are defined as usual. The informal reading of the modalities is: "Coalition C can choose ϕ", "It is permitted (/forbidden/obligated) for coalition C to choose ϕ", "It is rational for coalition C to choose ϕ".

[2] One interesting question is whether given any dynamic effectivity function and preference relation (with the above defined properties) we can always find a coalitionally dominated action (and hence a Pareto Efficient interaction). The acquainted reader will have noticed the resemblance of this problem with that of nonemptiness of the Core [11]. We leave though to further work the analysis of this relation. In case there is none, we may consider a satisfactory notion of optimal choice - as done for instance by Horty [7] - that looks at the relation between the choices in the choice sets of each coalition.

3.2 Structures

Definition 8 (Models) *A* model *for our logic is a quadruple*

$$(W, E, \{\geq_i\}_{i\subseteq Agt}, V)$$

where:

- W *is a nonempty set of states;*
- $E : W \longrightarrow (2^{Agt} \longrightarrow 2^{2^W})$ *is a coherent effectivity function;*
- $\geq_i \subseteq W \times W$ *for each* $i \in Agt$*, is the preference relation;*
- $V : W \longrightarrow 2^{Prop}$ *is the valuation function.*

3.3 Semantics

The satisfaction relation of the formulas with respect to a pointed model M, w is defined as follows:

$$\begin{array}{rcl}
M, w \models p & \text{iff} & p \in V(w) \\
M, w \models \neg\phi & \text{iff} & M, w \not\models \phi \\
M, w \models \phi \wedge \psi & \text{iff} & M, w \models \phi \text{ and } M, w \models \psi \\
M, w \models [C]\phi & \text{iff} & [[\phi]]^M \in E(w)(C) \\
M, w \models [rational_C]\phi & \text{iff} & \forall X (X \rhd_{C,w} \;\Rightarrow\; X \subseteq [[\phi]]^M) \\
M, w \models P(C, \phi) & \text{iff} & \exists X \in E(w)(C) \text{ s.t. } X \in \overline{VIOL}_{C,w} \text{ and } X \subseteq [[\phi]]^M \\
M, w \models F(C, \phi) & \text{iff} & \forall X \in E(w)(C)(X \subseteq \underline{[[\phi]]^M} \Rightarrow X \in VIOL_{C,w}) \\
M, w \models O(C, \phi) & \text{iff} & \forall X \in E(w)(C)(X \in \overline{VIOL}_{C,w} \Rightarrow X \subseteq [[\phi]]^M)
\end{array}$$

In this definition, $[[\phi]]^M =_{def} \{w \in W \mid M, w \models \phi\}$.

The modality for coalitional ability is standard from Coalition Logic [13]. The modality for rational action requires for a proposition ϕ to be rational (wrt a coalition C in a given state w) that all undominated choices (for C in w) be in the extension of ϕ. This means that there is no safe choice for a coalition that does not make sure that ϕ will hold. Notice that it is still possible for a coalition to pursue a rational choice that may be socially not rational.

The deontic modalities are defined in terms of the coalitional abilities and preferences. A choice is permitted whenever it is safe, forbidden when it may be unsafe (i.e. when it contains an inefficient choice), and obligated when it is the only choice that is safe.

4 Discussion

The definition of strong permission does not allow for a permitted choice of an agent to be refined by the other agents towards a violation. In fact we define permission for ϕ as "a ϕ-choice guarantees safety from violation". A more standard diamond modality would say " doing ϕ is compatible with no violation". A "safety" definition of permission has also been studied in [18], [14], [10], [3].

4.1 Properties

It is now interesting to look at what we can say and what we cannot say within our system.

Some Validities	
1	$P(C,\phi) \rightarrow \neg O(C,\neg\phi)$
2	$F(C,\phi) \leftrightarrow \neg P(C,\phi)$
3	$P(C,\phi) \vee P(C,\psi) \rightarrow P(C,\phi \vee \psi)$
4	$O(C,\phi) \rightarrow ([C]\phi \rightarrow P(C,\phi))$
5	$[rational_C]\phi \wedge [rational_{Agt}]\neg\phi \rightarrow F(C,\phi)$
6	$O(C,\phi) \vee O(C,\psi) \rightarrow O(C,\phi \vee \psi)$
7	$O(C,\top)$
8	$F(C,\phi) \wedge F(C,\psi) \rightarrow F(C,\phi \wedge \psi)$
9	$[rational_C]\phi \wedge (\phi \rightarrow \psi) \rightarrow [rational_C]\psi$

Some non-Validities	
10	$\neg O(C,\neg\phi) \rightarrow P(C,\phi)$
11	$P(C,\phi \vee \psi) \rightarrow P(C,\phi) \vee P(C,\psi)$
12	$O(C,\phi) \leftrightarrow \neg O(C,\neg\phi)$
13	$[rational_C]\phi \leftrightarrow [rational_{Agt}]\phi$
14	$O(C,\phi) \rightarrow P(C,\phi)$
15	$O(C,\phi \vee \psi) \rightarrow O(C,\phi) \vee O(C,\psi)$

The first validity says that the presence of permission imposes the absence of contrasting obligations, but the converse is not necessarily true. The second that prohibition and permission are interdefinable. The third says that the permission of ϕ or the permission of ψ implies the permission of ϕ or ψ. The fourth that the obligation to choose ϕ for an agent plus the ability to do something entails the permission to carry out ϕ. The validity number 5 says that the presence of a safe state that is rational for the grand coalition of agents is a norm for every coalition, even in case of conflicting preferences, i.e. in case of conflict the interest of the grand coalition prevails. The sixth one that obligation for ϕ or obligation for ψ implies the obligation for ϕ or ψ. Validity 7 says that there are no empty normative systems. The next validity says that prohibition is conjunctive. The last validity says that rational moves are monotonic. This has interesting implications on the choices of the agents, since refraining, i.e. choosing the biggest possible outcome, is always rational.

It is also useful to look at the non-validities: Number 10 says that if an agent is not obliged to choose something then it is permitted to do the contrary. But of course an agent may not be able to do anything, or may be not able to refine the choices "until the optimal". The next non-validity says that a permission of choice is not equivalent to a choice of permission. Number 12 says that a coalition can be obliged to do contradictory choices. This situation happens when a coalition is powerless or optimality is not possible. The next non validity says that the rational action for a certain coalition does not necessarily coincide with that of the grand coalition. Number 14, that ought does not imply can. The last does not allow to detach specific obligation from obligatory choices.

4.1.1 Further Assumptions

Playability Our notion of agency is more general than that of game theory. In particular we assume that even the grand coalition of agents may not determine a precise outcome of the interaction.

This is due to the abandonment of the property of playability of the effectivity function, that requires, together with regularity, outcome monotonicity, coalition monotonicity that:

- $X \notin E(C)$ implies $\overline{X} \in E(\overline{C})$, that is any choice excluded to a coalition is possible for the rest of the agents (maximality);
- For all X_1, X_2, C_1, C_2 such that $C_1 \cap C_2 = \emptyset$, $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$ imply that $X_1 \cap X_2 \in E(C_1 \cup C_2)$ (superadditivity)

Playability is a very strong property but it is needed to talk about games. As proved in [13] [Theorem 2.27], strategic games correspond exactly to playable effectivity functions [3]. With playable effectivity functions, the grand coalition can determine the exact outcome of the game and the dynamic effectivity function for the grand coalition of agents is the same in any state.

$$M, w \models [Agt]\phi \Leftrightarrow M \models [Agt]\phi$$

Moreover the fact that preferences do not change, induces the following stronger invariance:

$$M, w \models [rational_{Agt}]\phi \Leftrightarrow M \models [rational_{Agt}]\phi$$

So not only is every outcome reachable, but any situation shares the same social optimality. Notice that this is independent of the solution concept we may consider.

Finite Domain Another interesting assumption can be made about the finitness of the domain of discourse. With finite W, for $\mathcal{C}$ being the class of our models, we have that

Proposition 7 $\models_{\mathcal{C}} [rational_{Agt}]\phi$
implies that there exists an efficient outcome in the class of coalition models $\mathcal{C}$

Another property is the following:

$$\models_{\mathcal{C}} [rational_{Agt}]\neg\phi \wedge [C]\phi \rightarrow F(C, \phi) \tag{REG}$$

which says that any coalition has to refrain from a choice that is against an optimal state independently of its own preferences. A corresponding propery for obligation is instead the following:

$$\models_{\mathcal{C}} [rational_{Agt}]\neg\phi \wedge [C]\neg\phi \rightarrow O(C, \neg\phi) \tag{REG'}$$

[3]The proof involves the definition of strategic game as a tuple $\langle N, \{\Sigma_i | i \in N\}, o, S\rangle$ where N is a set of players, each i being endowed with a set of strategies σ_i from Σ_i, an outcome function that returns the result of playing individual strategies at each of the states in S; the definition of α effectivity function for a nonempty strategic game G, $E^{\alpha}_{G} : \wp(N) \rightarrow \wp\wp(S)$ defined as follows: $X \in E^{\alpha}_{G} \exists \sigma_C \forall \sigma_{\overline{C}} o(\sigma_C; \sigma_{\overline{C}}) \in X$. The above mentioned theorem establishes that $E^{\alpha}_{G} = E$ in case E is playable and G is a nonempty strategic game.

Coalitionally optimal norms The logic can be extended to treat norms that do not lead to a socially optimal outcome, but a coalitionally optimal outcome. That is it is possible to construct a deontic logic that pursues the interests of a particular coalition, independently of the other players' welfare. This extension is related to the work of Kooi and Tamminga on conflicting moral codes [8].

We limit the description to the obligation operator, the others are straightforward.
$M, w \models O^{C'}(C, \phi)$ iff $\forall X(X \rhd_{C,w}$ and $X \in \overline{VIOL}_{C,C',w} \Rightarrow X \subseteq [[\phi]]^M)$
where $VIOL_{C,C',w}$ is a C violation towards C', with $C \subseteq C'$.
For this operator it holds that

$$\models_{\mathcal{C}} O^C(C, \phi) \leftrightarrow [rational_C]\phi$$

that is playing for oneself boils down to rational action, and

$$\models_{\mathcal{C}} O^{Agt}(C, \phi) \leftrightarrow O(C, \phi)$$

that is, with the new operator we can express our original obligation operator.

4.2 Example: Norms of Cooperation

To consider forbidden all non optimal choices may seem a very strong requirement. Nevertheless, take the example in Table 1.

It is interesting to notice how $VIOL$ is not equivalent to the situations that each player is forbidden to choose. This is due to the fact that each player can only refine the choices of the other players, but cannot determine alone the outcome of the game: a permitted choice cannot be refined by permitted choices towards an inefficient outcome. Moreover $M \models [R]\neg(T_R) \wedge [rational_{R,C}](T_R)$, that by (REG) allows to conclude $F(R, \neg(T_R))$.

No agent is in fact obligated not to lie, but only permitted. Why is it so? Because no agent can alone reach a singleton state that is only good. But of course as a coalition $\{R, C\}$ has the obligation to end up in the optimal state.

Prisoner Dilemma, but think also of Coordination Games, in which individual players cannot reach a socially optimal outcome, have rules that say something about how coalitions should choose. This indirectly says something about the coalitions that are necessary to achieve an optimal outcome, i.e. about the coalitions that should form.

4.3 Future Work

The work here described allows for several developments. Among the most interesting ones is the study of the relation between imposed outcomes and steady states that describe where the game will actually end up (i.e. Nash Solution, the Core etc.). Conversely another feature that is worth studying is those structures in which Pareto Efficiency is not always present. Agents will reckon some actions as optimal even though there is no social equilibrium that can ever be reached. One more feature concerns the possibility of an inconsistent normative system. Further work could be done looking at the factual obedience of the norm, and how a norm affects preferences of agents (see for instance the work in [15]).

5 Conclusion

In this paper we proposed a deontic logic for optimal social norms. We described the concept of social optimality, explicitly linking it with the economical concept of Pareto Efficiency. Moreover we generalized the notion of Pareto Efficiency to capture those strategic interactions in which even the grand coalition of agents is not able to achieve every outcome.

Technically we did not assume playability of the coalitional effectivity functions. It is an important question in itself to understand the class of interactions to which such effectivity function corresponds. On top of the notion of Optimality we constructed a deontic language to talk about a normative system resulting from the imposition of such norms. We analyzed the properties of the language and discussed in details various examples from game theory and social science.

References

[1] K. Arrow. *Social Choice and Individual Values.* Yale University Press, 1970.

[2] S. Borgo. Coalitions in action logic. In *IJCAI*, pages 1822–1827, 2007.

[3] J. Broersen. Modal action logics for reasoning about reactive systems. PhD-thesis Vrije Universiteit Amsterdam, 2003.

[4] J. Broersen, J.J.Ch.Meyer, R.Mastop, and P.Turrini. A deontic logic for socially optimal norms. In *Proceedings Ninth International Workshop on Deontic Logic in Computer Science*, pages –, 2008.

[5] J. Coleman. *Foundations of Social Theory.* Belknap Harvard, 1990.

[6] P. Gardenfors. Rights, games and social choice. *Nous*, 15:341–56, 1981.

[7] J. Horty. *Deontic Logic and Agency.* Oxford University Press, 2001.

[8] B. Kooi and A.Tamminga. Conflicting obligations in multi-agent deontic logic. In J.-J. C. M. Lou Goble, editor, *Deontic Logic and Artificial Normative Systems: 8th International Workshop on Deontic Logic in Computer Science (DEON 2006)*, pages 175–186. LNCS 4048, 2006.

[9] F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity.* ILLC Dissertation Series, 2008.

[10] R. Mastop. What can you do? imperative mood in semantic theory. PhD thesis, ILLC Dissertation Series, 2005.

[11] M. Osborne and A. Rubinstein. *A course in Game Theory.* The MIT Press, 1994.

[12] R. Parikh. Social software. *Synthese*, 132(3):187–211, 2002.

[13] M. Pauly. *Logic for Social Software.* ILLC Dissertation Series, 2001.

[14] J. van Benthem. Minimal deontic logics. *Bulletin of the Section of Logic*, 8(1):36–42, 1979.

[15] J. van Benthem and F.Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 14, 2004.

[16] G. von Wright. *The logic of preference.* Edimburgh University Press, 1963.

[17] M. Weber. *Economy and Society.* Edited by Guenther Roth and Claus Wittich. New York: Bedminister Press, 1968.

[18] G. V. Wright. Deontic logic and the theory of conditions. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 3–31, 1971.