Available online at www.sciencedirect.com
ScienceDirect
Journal of Applied Logic 6 (2008) 152–171
JOURNAL OF APPLIED LOGIC
www.elsevier.com/locate/jal

Substantive and procedural norms in normative multiagent systems

Guido Boella [a,*], Leendert van der Torre [b]

[a] *Dipartimento di Informatica, Università di Torino, Italy*
[b] *Computer Science and Communications, University of Luxembourg, Luxembourg*

Available online 29 June 2007

Abstract

Procedural norms are instrumental norms addressed to agents playing a role in the normative system, for example to motivate these role playing agents to recognize violations or to apply sanctions. Procedural norms have first been discussed in law, where they address legal practitioners such as legislators, lawyers and policemen, but they are discussed now too in normative multiagent systems to motivate software agents. Procedural norms aim to achieve the social order specified using regulative norms like obligations and permissions, and constitutive norms like counts-as obligations. In this paper we formalize procedural, regulative and constitutive norms using input/output logic enriched with an agent ontology and an abstraction hierarchy. We show how our formalization explains Castelfranchi's notion of mutual empowerment, stating that not only the agents playing a role in a normative system are empowered by the normative system, but the normative system itself is also empowered by the agents playing a role in it. In our terminology, the agents are not only institutionally empowered, but they are also delegated normative goals from the system. Together, institutional empowerment and normative goal delegation constitute a mechanism which we call delegation of power, where agents acting on behalf of the normative system become in charge of recognizing which institutional facts follow from brute facts.

Keywords: Deontic logic; Multiagent systems; Normative systems; Procedural norms

1. Introduction

The distinction between substantive and procedural norms is well known in legal theory [35], but it seems that thus far procedural norms have not been formalized in deontic logic or introduced in formal models of normative multiagent systems. Substantive norms define the legal relationships of people with other people and the state in terms of regulative and constitutive norms, where regulative norms are obligations, prohibitions and permissions, and constitutive norms define what counts as institutional facts in the normative system. Procedural norms are instrumental norms, addressed to agents playing roles in the normative system, aiming to achieve the social order specified in terms of substantive norms [35]. In the Italian law, for example, it is obligatory for an attorney to start a prosecution process when he comes to know about a crime (art. 326 of *Codice di procedura penale*). Procedural law encompasses legal rules governing the process for settlement of disputes (criminal and civil). Procedural and substantive law are complementary. Procedural law brings substantive law to life and enables rights and duties to be enforced and defended.

* Corresponding author.
E-mail addresses: guido@di.unito.it (G. Boella), leendert@vandertorre.com (L. van der Torre).

1570-8683/$ – see front matter
doi:10.1016/j.jal.2007.06.006

For example, procedural norms explain how a trial should be carried out and which are the duties, rights and powers of judges, lawyers and defendants.

Substantive norms are discussed not only in law, but also in normative multiagent systems like electronic institutions. They are multiagent systems together with normative systems in which agents on the one hand can decide whether to follow the explicitly represented norms, and on the other the normative systems specify how and in which extent the agents can modify the norms [17].

In this paper we formalize procedural, regulative and constitutive norms using the following formal methods:

1. Input/output logic to define the various kinds of norms as logical relations between brute facts, decisions or actions and institutional facts in a logical framework without committing to one particular deontic logic. Various logical properties of the norms we define in this paper follow from general results in input/output logic (e.g., soundness and completeness with respect to the semantics of input/output logic) and are not discussed in this paper.
2. A multiagent ontology to define the concepts used in the system, not only to define the role playing agents in the normative system, but also to define a normative system as an organization or a legal actor.
3. An abstraction hierarchy to compare our detailed model with more abstract models and formalisms in the literature. For example, most deontic logics abstract from procedural norms or even from norms at all, and some normative multiagent systems assume that norm enforcement is done by cooperating or controlled agents only.

A success criterion for formal models of procedural norms, is which aspects of Castelfranchi's notion of mutual empowerment they can explain. Mutual empowerment says that not only the agents playing a role in a normative system are empowered by the normative system, but the normative system itself is also empowered by the agents playing a role in it. In our model, the agents are not only institutionally empowered, but they are also delegated normative goals from the system. In other words, the role playing agents empower the normative system by accepting its normative goals. Together, institutional empowerment and normative goal delegation constitute a mechanism which we call delegation of power, where agents acting on behalf of the normative system become in charge of recognizing which institutional facts follow from brute facts.

The paper is organized as follows. In Section 2 we informally discuss the concepts and mechanisms formalized in our logical framework. We motivate the introduction of procedural norms and the notion of delegation of power by means of an example, we discuss the role of procedural norms in regulative and constitutive norms, we define delegation of power and we compare it to goal delegation and institutionalized power. In Section 3 we introduce and explain our methodology of the abstraction hierarchy in which we distinguish five levels of abstraction at which normative multiagent systems can be represented. In Section 4 we introduce the formal model explained by examples of regulative, constitutive and procedural norms at the different levels of abstraction.

2. Procedural norms and delegation of power

In this section we informally discuss the concepts and mechanisms formalized in our logical framework. Section 2.1 motivates the introduction of the notions of procedural norms and delegation of power in normative multiagent systems by means of an example. Section 2.2 discusses the role of procedural norms in regulative and constitutive norms. Section 2.3 discusses in which sense normative systems are empowered by agents playing a role in it. Section 2.4 defines delegation of this power and compares it to goal delegation and delegation of power.

2.1. Cars with a catalytic converter count as ecological vehicles

To illustrate the notion of procedural norms and delegation of power playing a central role in the logical framework, we start with an example about traffic norms involving one of the authors. It is used as a running example throughout this paper.

Due to increased levels of pollution, on some precisely defined days, only ecological vehicles are allowed in major towns of Italy. One author of this paper bought many years ago one of the first catalytic cars, and since cars with a catalytic converter count as ecological vehicles, he felt permitted to always go around by car. On some day, he was stopped by the police and fined for driving around a non-ecological vehicle, because the car was bought before the local law recognized catalytic cars as ecological vehicles. The police agreed that the car had a catalytic converter: they

could see it, the car worked only with unleaded fuel, both the manual and the licence of the car said it has a catalytic converter. However, there was a missing rubber stamp by some office declaring that the car counts as an ecological vehicle.

The problem is not simply that only catalytic cars bought after a certain date are considered as ecological, but that a catalytic car is not ecological unless an agent officially recognizes it as such. The police has no power to consider the car as ecological, the evidence notwithstanding. The policeman may even argue that it is not the case that cars with a catalytic converter count as ecological vehicles, but that cars with the right stamps count as ecological vehicles—though this ignores the reason why the constitutive rule was introduced. The moral of the story is that even if a brute fact is present and could allow the recognition of an institutional fact, the institutional fact is the result of the action of some agent who is empowered to make institutional facts official.

For the layman, as the unlucky driver of the example, it is natural to describe the normative system as specifying the fact that catalytic cars are ecological. However, this is true only at a coarse level of detail. If the details of the functioning of the normative systems are entered, then a more complex description is necessary and agents with the specific role of making true some institutional facts under some circumstances must be introduced.

There may be many reasons why an institutional fact is the result of action of an empowered agent, for example for separation of duties, because it is not easy to identify whether a car has a catalytic converter, or simply for representational efficiency. We do not consider these reasons further in this paper.

2.2. *Regulative, constitutive and procedural norms*

Regulative norms specify the ideal and varying degrees of sub-ideal behavior of a system by means of obligations, prohibitions and permissions. Deontic logic [3,44] considers logical relations among obligations and permissions and focuses on the description of the ideal or optimal situation to achieve, driven by representation problems expressed by the so-called deontic paradoxes, most notoriously the contrary-to-duty paradoxes, see, for example, [30,43]. If we formalize the running example using regulative norms, we might say that only with a rubber stamp, cars are permitted to drive on these special days. However, norms are defined referring to more abstract states of affairs than catalytic converters and rubber stamps, like being an ecological vehicle.

Constitutive norms are based on the notion of counts-as and are used to support regulative norms by introducing institutional facts in the representation of legal reality. The notion of counts-as introduced by Searle [41] has been interpreted in deontic logic in different ways and it seems to refer to different albeit related phenomena [28]. For example, Jones and Sergot [31] consider counts-as from the constitutive point of view. According to Jones and Sergot, the fact that A counts-as B in context C is read as a statement to the effect that A represents conditions for guaranteeing the applicability of particular classificatory categories. The counts-as guarantees the soundness of that inference, and enables "new" classifications which would otherwise not hold. An alternative view of the counts-as relation is proposed by Grossi et al. [27]: according to the classificatory perspective A counts-as B in context C is interpreted as: A is classified as B in context C. In other words, the occurrence of A is a sufficient condition, in context C, for the occurrence of B. Via counts-as statements, normative systems can establish the ontology they use in order to distribute obligations, rights, prohibitions, permissions, etc. In [10,11] we propose a different view of counts-as which focuses on the fact that counts-as often provides an abstraction mechanism in terms of institutional facts, allowing the regulative rules to refer to legal notions which abstract from details. In all of these approaches, we can formalize the running example with a constitutive norm that a rubber stamp counts as an ecological vehicle, and a regulative norm stating that ecological vehicles are permitted to drive on the special days.

However, at the level of abstraction of these approaches we cannot express that cars with a catalytical converter count as ecological vehicles, since the evidence presents a counterexample. As long as we consider only the perspective of the unfortunate author, it may seem that there is also no need to do so. Clearly the evidence shows that cars with a catalytical converter not necessarily count as ecological vehicles, and there does not seem to be a reason why we should want to represent this norm. However, at the level of the normative multiagent system, for example from the perspective of the designer, it makes a lot of sense to state that catalytical cars count as ecological vehicles, since this may hold in the social order the normative system aims to achieve. Moreover, it is easier to communicate a norms like this to citizens.

The role that agents have in enforcing the social order the normative system aims to by creating norms has been recognized in normative multiagent systems [11,13], and agents are considered which are in charge of sanctioning

violations on behalf of the normative system [6,7]. Moreover, obligations are associated with procedural norms which are instrumental—to use Hart [29]'s terminology—to distribute the tasks to agents like judges and policemen, who have to decide whether and how to fulfill them. In the example, there are procedural norms that cars with a catalytical converter should count as ecological vehicles, and that drivers of non-catalytic cars that drive on the given days should be fined. Thus, at least informally, procedural norms can be seen as nestings of regulative and constitutive norms.

A formal framework for procedural norms has to explain how regulative and constitutive norms are related. Usually, constitutive norms are used in normative multiagent systems to define intermediate concepts like marriage or money, because they enable a more efficient representation to relate brute facts with deontic ones [38]. Another relation between the two kinds of norms is that an obligation to see to some state of affairs can be defined as "the absence of the state of affairs counts as a violation", generalizing Anderson's well-known reduction of deontic logic to alethic modal logic. Therefore, constitutive rules enter in the definition of sanction based obligations, since the recognition of a violation results in an institutional fact. Finally, with procedural norms, some agents may be obliged to recognize violations, to sanction violations, or to recognize some facts as institutional facts. These actions are carried out by agents playing roles in the normative system like judges and policemen.

2.3. *Mutual empowerment*

The normative system is an immaterial entity which exists due to the collective acceptance by the agents of a community and has the purpose to coordinate their activity. Thus, it cannot act in the environment, but it can act only by means of representatives. For example, there must be an agent which counts behavior as a violation, and one that applies the sanction in case of a violation. However, once there are such agents playing a role in it, we may say that the system is empowered by these agents, in the sense that it can act via these agents to achieve its goals. Since the normative system also gives some power to act in the normative system to the role playing agents, Castelfranchi has called the relation between a normative system and its agents *mutual empowerment* [22]. Procedural norms play an essential role in mutual empowerment, since they are the mechanism to see to it that the empowered agents empower the normative system, and do not abuse their power.

To represent that obligations are used to achieve social order, obligations can be defined as goals of the normative system that a certain state of affairs is achieved and that, if it is not achieved, that situation is recognized as a violation and sanctioned. Note the asymmetry between considering something as a violation and sanctioning. Sanctions can create new obligations like paying a fine, but they can also be physical actions like putting into jail, while a violation has always an institutional character. So, while a sanction of the latter type like putting into jail can be directly performed by a policeman, the recognition of a violation, and sanctions of the former type, can only be performed indirectly by means of some action which counts as the recognition of a violation, e.g., a trial by a judge which establishes that a violation happens and creates new obligations the violator is subject to.

Moreover, the same idea that norms can be defined as goals of the normative system can be used for constitutive norms. However, despite the fact that an action of an agent is necessary to create the institutional fact in case of ecological cars and in case of violations, there is an apparent asymmetry between constitutive rules and regulative ones. These cases can be modeled by a counts-as relation between the action of an agent (putting a stamp on the car licence or recognizing a violation) and the institutional fact (being an ecological vehicle or having violated an obligation), rather than by a direct counts-as relation between the brute facts and the institutional facts. But at first sight the two cases also have a difference: the recognition of a violation is wanted by the normative system to achieve its social order. In this case besides the counts-as rule between the action and the recognition as a violation there is also the goal of the normative system that this recognition contributes to the social order. As illustrated in the running example, in many circumstances facts which in principle should be considered as institutional facts are not recognized as such. In such circumstances, considering a fact as an institutional fact may depend on the action of some agent who is the representative of the normative system: we say that this agent has been *delegated the power* do recognize the fact as an institutional fact. We can see this as an instance of resource bounded reasoning: the assumption made above on constitutive rules, even if useful in some circumstances, is not realistic in all circumstances.

In other words, as a first step to explain why and how procedural norms are created, we should consider the motivational aspects behind constitutive norms required by agent theory. Constitutive norms are modeled as counts-as conditionals which allow to infer which institutional facts follow from brute facts and from existing institutional facts. E.g., a car counts as a vehicle for the traffic law. None of the models discussed in the previous section considers this

issue. The inference from facts to institutional facts is considered as automatic, i.e., it is assumed not to need any agent or resource to perform it. Agent theory, instead, considers also the resources needed to perform inferences, since it is aware that agents are usually resources bounded and that inferences have a cost too. Calculating the consequences following from some premises has a cost which must be traded off against the benefit of making the inferences. Thus in Georgeff and Ingrand [24] inferences are considered as actions which are planned and subject to decision processes as any other action: there must be an agent taking the decision about which inference to perform and executing the inference.

This mechanism allows also to give flexibility to the law, since the law cannot specify all possible cases in advance both as concerns regulative and constitutive rules. In this way, the decision to apply a constitutive rule is performed by an agent who can evaluate carefully the circumstances.

2.4. Delegation of power

Delegation of power involves two different phenomena, which are at first sight unrelated, institutional empowerment and goal delegation:

Institutional empowerment: how an action of some agent can establish an institutional fact.

Goal delegation: how this agent can carry out a goal of the delegator that the institutional fact holds in certain situations, and how it can be motivated to perform the action which establishes the institutional fact.

Empowerment means that an agent is made in the condition to satisfy the goal of some agent. In general, it has no institutional character. However, institutional empowerment is by nature an institutional phenomenon which is based on the counts-as relation: an agent is empowered to perform an institutional action—a kind of institutional fact—if some of its actions counts as the institutional action. The institutional action has some effects on the institution which empowers the agent. For example, a director can commit by means of his signature his institution to purchase some goods. Thus it is essentially related to counts-as rules, albeit restricted to actions of agents. Consider as a paradigmatic case the work by Jones and Sergot [31]. However, it is not sufficient to explain mutual empowerment either, since it does not explain how it empowers the normative system. In particular, the empowerment of the normative system is not institutional, but it consists in the actions which agents carry out on behalf of the normative system, which, by itself, cannot act.

According to Castelfranchi [21], goal delegation is relying on another agent for the achievement of one own's goal: "in delegation an agent A needs or likes an action of another agent B and includes it in its own plan. In other words, A is trying to achieve some of its goals through B's behaviors or actions; thus A has the goal that B performs a given action/behavior". Goal delegation by itself is not sufficient to explain mutual empowerment, since it is not an institutional phenomenon but a basic capability of agents which enables them to interact with each other.

Bottazzi and Ferrario [18] argue that the two phenomena are related, because an agent which is institutionally empowered, is also delegated the goal of the institution of making true an institutional fact by exercising his power in the specified situations. The connection between goal delegation and institutional empowerment is not a necessary one. There are situations where an agent is delegated but not institutionally empowered. For example, the agent in charge for sanctioning an obligation is delegated the goal of sanctioning, but there is no need of institutional powers in case of physical sanctions: the agent's physical powers are sufficient, for instance, to put the violator into jail. Viceversa, the law institutionally empowers agents to stipulate private contracts which have the force of law, without being delegated by the law to do so, since contracting agents act for their own sake [11].

This connection, which we call *delegation of power*, can be used to explain the running example. For the normative system, catalytic cars have to be considered as ecological vehicles. There are three possibilities: first, recognizing all catalytic cars as ecological vehicles by means of a constitutive norm. This solution, however, does not consider the actual performance of the inference and the possible costs related to it. Second, the normative system can rely on some agent to recognize catalytic cars as ecological vehicles. As said above, this can be done by means of a counts-as relation between an action of an agent and its effects. This solution, however, fails to account for the motivations that the agent should have to perform the action of recognizing ecological vehicles as such. Third, also a goal of the normative system is added to motivate his action: there is an agent who has the institutional power to recognize cars

as ecological vehicles and the normative system has delegated it the goal that it does so in order to motivate it. If this motivation is not sufficient since the agent is not cooperative, an obligation to achieve this goal of the normative system must be added.

3. Level of abstractions in the definition of norms

Normative Multiagent Systems can be described at different levels of abstraction, using definitions of obligations and counts-as relations which differ at their level of detail. We identify five different levels of abstraction, which are suited for different applications, depending on the role attributed to the system infrastructure or on specialized agents. The abstraction dimension is the detail at which we consider agents acting for the normative system: at the higher abstraction level agents have no role, at the next abstraction level only actions of the normative system are considered but agents are not considered; then at more concrete level, where agents are in charge of the actual functioning of the normative system concerning regulative and constitutive rules, delegation of power enters the picture; finally, at the more concrete level of abstraction, procedural norms are introduced to motivate the agents playing roles in the system. In the next section we will provide the formal definition of obligations and counts-as relations at the different levels.

1. The higher level abstracts from the fact that violations, sanctions and institutional facts are the result of the action that some agent desires and decides to perform. The obligations are defined as in Anderson's reduction [2], in the sense that the recognition of the violation and the sanction follow necessarily from the violation. This abstraction level for regulative rules is adopted also by [4,36] and we use it in [15]. For constitutive rules, up to our knowledge, this is the only level considered. The responsibility of monitoring, sanctioning and making counts-as inferences is completely delegated to the infrastructure.
2. At the second level the normative system is in some way personified and is attributed mental attitudes, thus, the recognition of violations and sanctions are wanted by the normative system itself. If obligations are goals of the normative system, the institutional facts follow from the beliefs of the normative system: they are still logical consequences of facts, but in the beliefs of the normative system. This is the solution adopted in [10].
3. The third level not only abstracts from the role of agents in the normative system, but the recognition of violations and sanctions are considered as the actions of the normative system itself. We adopt this level of representation for regulative norms in [11,13]. Analogously, institutional facts follow from actions of the normative system: they are not anymore logical consequences of beliefs, but consequences of decisions of the normative systems which are traded-off against other decisions. They, thus, do not follow automatically, since the normative system can take a different decision due to conflicts with other goals or to lack of resources. The normative system can decide to enforce norms or not in specific situations, thus allowing interpretation of them.
4. The fourth level takes into account the actions of the agents acting on behalf of the normative system. Concerning regulative norms, some agents are delegated the goal to sanction violations and the goal and power of recognizing violations. I.e., they are delegated the power to do so. Concerning constitutive norms, the agents are delegated the goal to recognize some facts as institutional facts and the power to do so by means of their actions. I.e., they are delegated the power to do so. The problem of agents recognizing violations has been partially addressed in [6], but the recognition action was considered as a physical action like the sanction. In this paper we add the counts-as relation to the recognition of violations.
5. The fifth level introduces procedural norms to motivate the agents playing roles in the system. These norms oblige agents playing roles to achieve the goals of the normative system. Concerning regulative norms, some agents are obliged to achieve the delegated goal to sanction violations and the goal and power of recognizing violations. Concerning constitutive norms, the agents are obliged to achieve the goal to recognize some facts as institutional facts by exercising the delegated powers. In the Italian law, for example, it is obligatory for an attorney to start a prosecution process when he comes to know about a crime (art. 326 of *Codice di procedura penale*).

Our logical framework of the game-theoretic approach to normative multiagent systems is based on the so-called agent metaphor: social entities like normative systems can be conceptualized as agents by attributing them mental attitudes like beliefs and goals. The agent metaphor underlying our framework is based on cognitive motivations [9]. Beliefs model the constitutive rules of the normative system, while goals model regulative rules. Thus, in the normative system the interaction between constitutive and regulative rules is the same as the interaction of beliefs and goals in

an agent. However, differently from a real agent, the normative system is a socially constructed agent. It exists only because of the collective acceptance by all the agents and, thus, it cannot act in the world. Its actions are carried out by agents playing roles in the normative system, like legislators, judges and policemen. It is a social construction used to coordinate the behavior of agents.

Obligations are not only modeled as goals of the normative system, but they are also associated with the instrumental goals that the behavior of the addressee of the norms is considered as a violation and that the violation is sanctioned. Considering something as a violation and sanctioning are actions which can be executed by the normative system itself, or, at a more concrete level of detail, by agents playing roles in it.

The counts-as relation in our model is modeled as a conditional belief of the normative system to provide an abstraction of reality in terms of institutional facts. Regulative norms can refer to this level, thus decoupling them from the details of reality. For example, the institutional fact that traffic lights are red must be distinguished from the brute fact that red light bulbs in the traffic lights are on: in the extreme case the institutional fact can be true even if all the red bulbs are broken. In the following section, we consider how counts-as can be used to define delegation of power. Counts-as relations are not used in this case to directly connect brute facts to institutional facts, but only to express the institutional effects of actions of agents empowered by the normative system. This is analogous to the act of signing of the director counts as the commitment of the institution he belongs to. Moreover, constitutive rules specify also how the normative system can change [11]. Since it is a socially constructed agent, it cannot directly change itself, but it relies on the actions of agents playing roles in it, like legislators, which count as changes of the system.

4. The formal model

The definition of the agents is inspired by the rule based BOID architecture [20], though in our theory, and in contrast to BOID, obligations are not taken as primitive concepts. Beliefs, desires and goals are represented by conditional rules rather than in a modal framework. Intentions have been introduced as a form of bounded rationality: since an agent has not enough resources to make the optimal decision at each moment, he maintains its previous choices. In this paper we consider only one decision, so we do not need to introduce intentions to model decisions which persist over time.

4.1. Input/output logic

To represent conditional mental attitudes we take a simplified version of the input/output logics introduced in [33,34]. Though the development of input/output logic has been motivated by the logic of norms, the same logic can be used for other conditionals like conditional beliefs and conditional goals—which explains the more general name of the formal system. Moreover, Bochman [5] also illustrates how the same logic is used for causal reasoning and various non-monotonic reasoning formalisms.

A rule set is a set of ordered pairs $P \to q$. For each such pair, the body P is thought of as an input, representing some condition or situation, and the head q is thought of as an output, representing what the rule tells us to be believed, desirable, obligatory or whatever in that situation. In this paper, to keep the formal exposition simple, input and output are respectively a set of literals and a literal.

Definition 1 *(Input/output logic).* Let X be a set of propositional variables, the set of literals built from X, written as $Lit(X)$, is $X \cup \{\neg x \mid x \in X\}$, and the set of rules built from X, written as $Rul(X) = 2^{Lit(X)} \times Lit(X)$, is the set of pairs of a set of literals built from X and a literal built from X, written as $\{l_1, \ldots, l_n\} \to l$. We also write $l_1 \wedge \cdots \wedge l_n \to l$ and when $n = 0$ we write $\top \to l$. For $x \in X$ we write $\sim x$ for $\neg x$ and $\sim(\neg x)$ for x. Moreover, let Q be a set of pointers to rules and $RD : Q \to Rul(X)$ is a total function from the pointers to the set of rules built from X.

Let $S = RD(Q)$ be a set of rules $\{P_1 \to q_1, \ldots, P_n \to q_n\}$, and consider the following proof rules strengthening of the input (SI), disjunction of the input (OR), cumulative transitivity (CT) and Identity (Id) defined as follows:

$$\frac{p \to r}{p \wedge q \to r} SI \quad \frac{p \wedge q \to r,\ p \wedge \neg q \to r}{p \to r} OR \quad \frac{p \to q,\ p \wedge q \to r}{p \to r} CT \quad \frac{}{p \to p} Id.$$

The following output operators are defined as closure operators on the set S using the rules above.

out_1: SI (simple-minded output) out_3: $SI + CT$ (reusable output)

out_2: $SI + OR$ (basic output) out_4: $SI + OR + CT$ (basic reusable output).

Moreover, the following four throughput operators are defined as closure operators on the set S. out_i^+: $out_i + Id$ (throughput). We write $out(Q)$ for any of these output operations and $out^+(Q)$ for any of these throughput operations. We also write $l \in out(Q, L)$ iff $L \to l \in out(Q)$, and $l \in out^+(Q, L)$ iff $L \to l \in out^+(Q)$.

Example 1. Given $RD(Q) = \{a \to x, x \to z\}$ the output of Q contains $x \wedge a \to z$ using the rule SI. Using also the CT rule, the output contains $a \to z$. $a \to a$ follows only if there is the Id rule.

A technical reason to distinguish pointers from rules is to facilitate the description of the priority ordering we introduce in the following definition.

The notorious contrary-to-duty paradoxes such as Chisholm's and Forrester's paradox have led to the use of constraints in input/output logics [34]. The strategy is to adapt a technique that is well known in the logic of belief change—cut back the set of norms to just below the threshold of making the current situation inconsistent.

Definition 2 *(Constraints).* Let $\geqslant : 2^Q \times 2^Q$ be a transitive and reflexive partial relation on the powerset of the pointers to rules containing at least the subset relation and $RD : Q \to Rul(X)$ a function from the pointers to the set of rules. Moreover, let out be an input/output logic:

- $maxfamily(Q, P)$ is the set of $\subseteq$-maximal subsets Q' of Q such that $out(Q', P) \cup P$ is consistent.
- $preffamily(Q, P, \geqslant)$ is the set of $\geqslant$-maximal elements of $maxfamily(Q, P)$.
- $outfamily(Q, P, \geqslant)$ is the output under the elements of $preffamily$, i.e., $\{out(Q', P) \mid Q' \in preffamily(Q, P, \geqslant)\}$.
- $P \to x \in out_\cup(Q, \geqslant)$ iff $x \in \cup outfamily(Q, P, \geqslant)$, $P \to x \in out_\cap(Q, \geqslant)$ iff $x \in \cap outfamily(Q, P, \geqslant)$.

Example 2. Let $RD(\{a, b, c\}) = \{a = (\top \to m), b = (p \to n), c = (o \to \neg m)\}$, $\{b, c\} > \{a, b\} > \{a, c\}$, where by $A > B$ we mean as usual $A \geqslant B$ and $B \not\geqslant A$.

$$maxfamily(Q, \{o\}) = \{\{a, b\}, \{b, c\}\},$$
$$preffamily(Q, \{o\}, \geqslant) = \{\{b, c\}\},$$
$$outfamily(Q, \{o\}, \geqslant) = \{\{\neg m\}\}.$$

The *maxfamily* includes the sets of applicable compatible pointers to rules together with all non-applicable ones: e.g., the output of $\{a, c\}$ in the context $\{o\}$ is not consistent. Finally $\{a\}$ is not in *maxfamily* since it is not maximal, we can add the non-applicable rule b. Then *preffamily* is the preferred set $\{b, c\}$ according to the ordering on set of rules above. The set *outfamily* is composed by the consequences of applying the rules $\{b, c\}$ which are applicable in o (c): $\neg m$.

The semantics of input/output logic given by Makinson and van der Torre [33] is an operational semantics, which characterizes the output as a function of the input and the set of norms. Makinson and van der Torre illustrate how to recapture input/output logic in modal logic, and thus give it a classical possible worlds semantics. Bochman [5] illustrates how the operational semantics of input/output logic can be rephrased as a bimodel semantics, in which a model of a set of conditionals is a pair of partial models from the base logic (here, propositional logic). Due to space limitations we have to be brief on details with respect to input/output logics, see [33,34] for the semantics of input/output logics, further details on its proof theory, its possible translation to modal logic, alternative constraints, and examples.

4.2. Multiagent systems

We assume that the base language contains Boolean variables and logical connectives. The variables are either *decision variables* of an agent, which represent the agent's actions and whose truth value is directly determined by it, or *parameters*, which describe both the state of the world and *institutional facts*, and whose truth value can only be

determined indirectly. Our terminology is borrowed from Lang et al. [32] and is used in discrete event systems, and many formalisms in operations research.

Given the same set of mental attitudes, agents reason and act differently: when facing a conflict among their motivations and beliefs, agents may prefer to fulfill distinct goals and desires. We express these agent characteristics by a priority relation on the mental attitudes which encode how the agent resolves its conflicts [20]. The priority relation is defined on the powerset of the mental attitudes such that a wide range of characteristics can be described, including social agents that take the desires or goals of other agents into account. The priority relation contains at least the subset-relation which expresses a kind of independence among the motivations.

Background knowledge is formalized by a set of effects E represented by rules.

Definition 3 *(Agent set).* An agent set is a tuple $\langle A, X, B, D, G, AD, E, \geqslant, \geqslant_E\rangle$, where:

- The agents A, propositional variables X, agent beliefs B, desires D, goals G, and effects E are six finite disjoint sets. B, D, G are sets of mental attitudes. We write $M = D \cup G$ for the motivations defined as the union of the desires and goals. The set of effects E represents the background knowledge of all agents.
- An agent description $AD: A \to 2^{X \cup B \cup M}$ is a total function that maps each agent to sets of variables (its decision variables), beliefs, desires and goals, but that does not necessarily assign each variable to at least one agent. For each agent $b \in A$, we write X_b for $X \cap AD(b)$, and B_b for $B \cap AD(b)$, D_b for $D \cap AD(b)$, etc. We write parameters $P = X \setminus \bigcup_{b \in A} X_b$.
- A priority relation $\geqslant: A \to 2^B \times 2^B \cup 2^M \times 2^M$ is a function from agents to a transitive and reflexive partial relation on the powerset of the motivations containing at least the subset relation. We write $\geqslant_b$ for $\geqslant(b)$.
- A priority relation $\geqslant_E: 2^E \times 2^E$ is a transitive and reflexive partial relation on the powerset of effects containing at least the subset relation.

Example 3. $A = \{a\}$, $X_a = \{drive\}$, $P = \{s, catalytic\}$, $D_a = \{d_1, d_2\}$, $\geqslant_a = \{d_2\} \geqslant \{d_1\}$. There is a single agent, agent a, who can drive a car. Moreover, it can be sanctioned and the car can be catalytic. It has two desires, one to drive (d_1), another one not to be sanctioned (d_2). The second desire is more important.

In a multiagent system, beliefs, desires, goals and effects are abstract concepts which are described by rules built from literals.

Definition 4 *(Multiagent system).* A multiagent system, *NMAS*, is a tuple $\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E\rangle$, where $\langle A, X, B, D, G, AD, E, \geqslant, \geqslant_E\rangle$ is an agent set, and the rule description $RD: (B \cup M \cup E) \to Rul(X)$ is a total function from the sets of beliefs, desires and goals, and effects to the set of rules built from X. For a set of pointers $S \subseteq B \cup M \cup E$, we write $RD(S) = \{RD(q) \mid q \in S\}$.

Example 4 *(Continued).* $RD(d_1) = \top \to drive$, $RD(d_2) = \top \to \neg s$.

In the description of the normative system, we do not introduce norms explicitly, but we represent several concepts which are illustrated in the following sections. Institutional facts (I) represent legal abstract categories which depend on the beliefs of the normative system and have no direct counterpart in the world. $F = X \setminus I$ are what Searle calls "brute facts": physical facts like the actions of the agents and their effects. $V(x, a)$ represents the decision of agent $\mathbf{n}$ that recognizes x as a violation by agent a. The goal distribution $GD(a) \subseteq G_{\mathbf{n}}$ represents the goals of agent $\mathbf{n}$ the agent a is responsible for.

Definition 5 *(Normative system).* A normative multiagent system, written as *NMAS*, is a tuple

$$\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E, \mathbf{n}, I, V, GD\rangle$$

where the tuple $\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E\rangle$ is a multiagent system, and

- The normative system $\mathbf{n} \in A$ is an agent.
- The institutional facts $I \subseteq P$ are a subset of the parameters.

- The norm description $V : Lit(X) \times A \to X_{\mathbf{n}} \cup P$ is a function from the literals and the agents to the decision variables of the normative system and the parameters.
- The goal distribution $GD : A \to 2^{G_{\mathbf{n}}}$ is a function from the agents to the powerset of the goals of the normative system, such that if $L \to l \in RD(GD(a))$, then $l \in Lit(X_a \cup P)$.

Agent **n** is a normative system with the goal that non-catalytic cars are not driven, i.e., with the aim to achieve the social order in which such cars are not driven.

Example 5 *(Continued).* There is agent **n**, representing the normative system.

$$P = \{s, V(drive, a), catalytic\}, \qquad D_{\mathbf{n}} = G_{\mathbf{n}} = \{g_1\},$$
$$RD(g_1) = \{\neg catalytic \to \neg drive\}, \qquad GD(a) = \{g_1\}.$$

The parameter $V(drive, a)$ represents the fact that the normative system considers a violation agent a's action of driving. It has the goal that non-ecological vehicles should not be driven by a and it has distributed this goal to agent a.

In the following, we use an input/output logic *out* to define whether a desire or goal implies another one and to define the application of a set of belief rules to a set of literals; in both cases we use the out_3 operation since it has the desired logical property of not satisfying identity.

We now define obligations and the counts-as relation at the five levels of abstraction.

4.3. First level of abstraction (highest)

Regulative norms are conditional obligations with an associated sanction. At the higher level of abstraction, the definition contains three clauses: the first two clauses state that recognitions of violations and sanctions are a consequence of the behavior of agent a, as it is represented by the background knowledge rules E. For an obligation to be effective, the third clause states that the sanction must be disliked by its addressee.

Definition 6 *(Obligation (level 1)).* Let *NMAS* be a normative multiagent system

$$\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E, \mathbf{n}, I, V, GD\rangle.$$

Agent $a \in A$ is *obliged* to see to it that $x \in Lit(X_a \cup P)$ with $V(\sim x, a) \in Lit(P)$ and sanction $s \in Lit(P)$ in context $Y \subseteq Lit(X)$ in *NMAS*, written as $NMAS \models O^1_{a\mathbf{n}}(x, s|Y)$, if and only if:

1. $Y \cup \{\sim x\} \to V(\sim x, a) \in out(E, \geqslant_E)$: if Y and x is false, then it follows that $\sim x$ is a violation by agent a.
2. $Y \cup \{V(\sim x, a)\} \to s \in out(E, \geqslant_E)$: if Y and there is a violation by agent a, then it is sanctioned.
3. $Y \to \sim s \in out(D_a, \geqslant_a)$: if Y, then agent a desires $\sim s$, which expresses that it does not like to be sanctioned.

Example 6. Let: $E = \{e_1, e_2\}$, $D_a = \{d_2\}$,

$$RD(e_1) = \{\neg catalytic, drive\} \to V(drive, a),$$
$$RD(e_2) = \{\neg catalytic, V(drive, a)\} \to s,$$
$$RD(d_2) = \neg catalytic \to \sim s,$$

$NMAS \models O^1_{a\mathbf{n}}(\neg drive, s \mid \neg catalytic)$, since:

1. $\{\neg catalytic, drive\} \to V(drive, a) \in out(E, \geqslant_E)$,
2. $\{\neg catalytic, V(drive, a)\} \to s \in out(E, \geqslant_E)$,
3. $\neg catalytic \to \sim s \in out(D_a, \geqslant_a)$.

To determine what the agent a will do, we can define a qualitative decision problem that considers the consequences of its decisions, and chooses the decision which achieves the goals with the highest priority. Here it will choose not to drive a non-catalytic car, because it does not want to be sanctioned.

Constitutive norms introduce new abstract categories of existing facts and entities, called institutional facts. In [10] we formalize the counts-as conditional as a belief rule of the normative system **n**. However, since at the first level of abstraction we do not consider yet the normative system, counts-as conditionals are part of the background knowledge rules E.

The condition x of the rule is a variable which can be an action of an agent, a brute fact or an institutional fact. So, the counts-as relation can be iteratively applied.

Definition 7 *(Counts-as relation (level 1)).* Let *NMAS* be a normative multiagent system

$$\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E, \mathbf{n}, I, V, GD\rangle.$$

A literal $x \in Lit(X)$ counts-as $y \in Lit(I)$ in context $C \subseteq Lit(X)$, $NMAS \models counts\text{-}as^1_{\mathbf{n}}(x, y|C)$, iff $C \cup \{x\} \rightarrow y \in out(E, \geqslant_E)$: y is a consequence of C and x.

Example 7. $P \setminus I = \{catalytic\}$, $I = \{eco\}$, $X_a = \{drive\}$, $E = \{b_1\}$, $RD(b_1) = catalytic \rightarrow eco$.

Consequently, $NMAS \models counts\text{-}as^1_{\mathbf{n}}(catalytic, eco|\top)$. This formalizes that for the normative system a catalytic car counts as an ecological vehicle. The presence of the catalytic converter is a physical "brute" fact, while being an ecological vehicle is an institutional fact. In situation $S = \{catalytic\}$, given E we have that the consequences of the constitutive norms are $out(E, S, \geqslant_E) = \{eco\}$ (since out_3 does not include Id).

Note that the institutional facts can appear in the conditions of regulative norms.

Example 8. A regulative norm which forbids driving non-catalytic cars can refer to the abstract concept of ecological vehicle rather than to catalytic converters:

$$NMAS \models O^1_{a\mathbf{n}}(\neg drive, s|\neg eco)$$

To define the optimal decision of an agent in a normative system incorporating both the regulative and the constitutive norm, the consequences of its decisions first take into account which institutional effects follow from the brute facts, and thereafter the decisions of the normative system.

As the system evolves, new cases can be added to the notion of ecological vehicle by means of new constitutive norms, without changing the regulative norms about it. E.g., if a car has fuel cells, then it is an ecological vehicle: $fuelcell \rightarrow eco \in RD(E)$.

4.4. Second level of abstraction

At the second level of abstraction we introduce the normative system as an agent described by mental attitudes like beliefs and goals. We do not consider, however, how its goals are achieved in the multiagent system. Thus, it represents only an abstract specification of the ideal behavior of the system.

The first and central clause of our definition of obligation defines obligations of agents as goals of the normative system, following the "your wish is my command" metaphor. It says that the obligation is implied by the desires of the normative system **n**, implied by the goals of agent **n**, and it has been distributed by agent **n** to the agent. The latter two steps are represented by $out(GD(a), \geqslant_{\mathbf{n}})$.

The second and third clause can be read as the normative system has the goal that the absence of p is considered as a violation. The third clause says that the agent desires that there are no violations, which is stronger than that it does not desire violations, as would be expressed by $\top \rightarrow V(\sim x, a) \notin out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}})$.

The fourth and fifth clause relate violations to sanctions. The fifth clause says that the normative system is motivated not to sanction as long as their is no violation, because otherwise the norm would have no effect. Finally, for the same reason the last clause says that the agent does not like the sanction.

Definition 8 *(Obligation (level 2)).* Let *NMAS* be a normative multiagent system

$$\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E, \mathbf{n}, I, V, GD\rangle.$$

Agent $a \in A$ is *obliged* to see to it that $x \in Lit(X_a \cup P)$ with $V(\sim x, a) \in Lit(P)$ and sanction $s \in Lit(P)$ in context $Y \subseteq Lit(X)$ in *NMAS*, written as $NMAS \models O^2_{a\mathbf{n}}(x, s|Y)$, if and only if:

1. $Y \rightarrow x \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(GD(a), \geqslant_{\mathbf{n}})$: if Y holds then agent **n** desires and has as a goal that x, and this goal has been distributed to agent a.
2. $Y \cup \{\sim x\} \rightarrow V(\sim x, a) \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geqslant_{\mathbf{n}})$: if Y holds and $\sim x$, then agent **n** has the goal and the desire $V(\sim x, a)$: that it is recognized as a violation by agent a.
3. $\top \rightarrow \neg V(\sim x, a) \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}})$: agent **n** desires that there are no violations.
4. $Y \cup \{V(\sim x, a)\} \rightarrow s \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geqslant_{\mathbf{n}})$: if Y holds and agent **n** decides $V(\sim x, a)$, then agent **n** desires and has as a goal that agent a is sanctioned.
5. $Y \rightarrow \sim s \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}})$: if Y holds, then agent **n** desires not to sanction. This desire of the normative system expresses that it only sanctions in case of violation.
6. $Y \rightarrow \sim s \in out(D_a, \geqslant_a)$: if Y holds, then agent a desires $\sim s$, which expresses that it does not like to be sanctioned.

The beliefs, desires and goals of the normative agent—defining the obligations—are not private mental states of an agent. Rather they are collectively attributed by the agents of the normative system to the normative agent: they have a public character, and, thus, which are the obligations of the normative system is a public information.

In [10] we formalize the counts-as conditional as a belief rule of the normative system **n**. At the second level we do the same, to separate brute facts and their relations E from institutional facts. Note that in our model the counts-as relation does not satisfy the identity rule. See [10] for a discussion of the motivations.

Definition 9 *(Counts-as relation (level 2)).* Let *NMAS* be a normative multiagent system

$$\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E, \mathbf{n}, I, V, GD\rangle.$$

A literal $x \in Lit(X)$ counts-as $y \in Lit(I)$ in context $C \subseteq Lit(X)$, $NMAS \models counts\text{-}as^1_{\mathbf{n}}(x, y|C)$, iff $C \cup \{x\} \rightarrow y \in out(B_{\mathbf{n}}, \geqslant_{\mathbf{n}})$: if agent **n** believes C and x then it believes y.

Example 9. $P \setminus I = \{catalytic\}$, $I = \{eco\}$, $X_a = \{drive\}$, $B_{\mathbf{n}} = \{b_1\}$, $RD(b_1) = catalytic \rightarrow eco$.

Consequently, $NMAS \models counts\text{-}as^1_{\mathbf{n}}(catalytic, eco|\top)$. This formalizes that for the normative system a catalytic car counts as an ecological vehicle. The presence of the catalytic converter is a physical "brute" fact, while being an ecological vehicle is an institutional fact. In situation $S = \{catalytic\}$, given E we have that the consequences of the constitutive norms are $out(B_{\mathbf{n}}, S, \geqslant_{\mathbf{n}}) = \{eco\}$ (and not *catalytic*, since out_3 does not include *Id*).

4.5. Third level of abstraction

At the next level of abstraction, also for constitutive rules actions of the normative systems are added in the definition of the obligations: the recognition of a violation and sanctions. Since the actions undergo a decision process, desires and goals of the normative system are added like at the previous level.

Definition 10 *(Obligation (level 3)).* $NMAS \models O^3_{a\mathbf{n}}(x, s|Y)$, iff $NMAS \models O^2_{a\mathbf{n}}(x, s|Y)$ and $s \in Lit(X_{\mathbf{n}})$ and $V(\sim x, a) \in Lit(X_{\mathbf{n}})$.

The rules in the definition of obligation are only motivations, and not beliefs, because a normative system may not recognize that a violation counts as such, or that it does not sanction it: it is up to its decision. Both the recognition of the violation and the application of the sanction are the result of autonomous decisions of the normative system that is modeled as an agent. In this way, regulative norms can be interpreted by the normative system, who takes a decision depending on the circumstances.

Example 10. Let: $G_{\mathbf{n}} = \{g_1, g_2, g_4\}$, $G_{\mathbf{n}} \cup \{g_3, d_2\} = D_{\mathbf{n}}$, $\{g_1\} = GD(a)$, $\{d_2\} = D_a$,

$$RD(g_2) = \{\neg catalytic, drive\} \rightarrow V(drive, a), \qquad RD(g_3) = \top \rightarrow \neg V(drive, a),$$
$$RD(g_4) = \{\neg catalytic, V(drive, a)\} \rightarrow s,$$

$NMAS \models O^3_{a\mathbf{n}}(\neg drive, s \mid \neg catalytic)$, since:

1. $\neg catalytic \rightarrow \neg drive \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(GD(a), \geqslant_{\mathbf{n}})$,
2. $\{\neg catalytic, drive\} \rightarrow V(drive, a) \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geqslant_{\mathbf{n}})$,
3. $\top \rightarrow \neg V(drive, a) \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}})$,
4. $\{\neg catalytic, V(drive, a)\} \rightarrow s \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geqslant_{\mathbf{n}})$,
5. $\neg catalytic \rightarrow \sim s \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}})$,
6. $\neg catalytic \rightarrow \sim s \in out(D_a, \geqslant_a)$.

It is only at this level of abstraction that we formalize the game-theoretic approach to normative systems, in the sense that to determine the optimal decision of an agent, we have to take the response of the normative system into account by recursively modeling it [25]. Moreover, we can also play more complex games, for example one in which the normative system decides which norm to introduce by recursively modeling agents recursively modeling the normative system again, and so on.

At the third level of abstraction, we introduce in the model the actions of the normative system. The beliefs of the normative system are restricted to the connections between actions and the consequences of these actions for the normative system. The normative system has the desire and goal that the institutional fact y holds if the fact x holds in context C. The normative system believes that to make y true it has to perform an action z. Thus the fact x holding in context C is not sufficient for the institutional fact y to be true: a decision is necessary also to do z by the normative system.

Definition 11 *(Counts-as relation (level 3)).* Let *NMAS* be a normative multiagent system

$$\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E, \mathbf{n}, I, V, GD\rangle.$$

A literal $x \in Lit(X)$ counts-as $y \in Lit(I)$ in context $C \subseteq Lit(X)$, $NMAS \models counts\text{-}as^3_{\mathbf{n}}(x, y|C)$, iff:

1. $C \cup \{x\} \rightarrow y \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geqslant_{\mathbf{n}})$: it is a desire and goal of the normative system that in context C the fact x is considered as the institutional fact y.
2. $\exists z \in X_{\mathbf{n}}$ such that $C \cup \{z\} \rightarrow y \in out(B_{\mathbf{n}}, \geqslant_{\mathbf{n}})$: there exists an action z of the normative system $\mathbf{n}$ such that if it decides z in context C then it believes that the institutional fact y follows (i.e., $counts\text{-}as^2_{\mathbf{n}}(z, y|C)$ at the second level of abstraction).

Example 11. $P \setminus I = \{catalytic\}$, $I = \{eco\}$, $X_a = \{drive\}$, $X_{\mathbf{n}} = \{stamp\}$,

$$D_{\mathbf{n}} = G_{\mathbf{n}} = \{d_3\}, \qquad RD(d_3) = catalytic \rightarrow eco,$$
$$B_{\mathbf{n}} = \{b_1\}, \qquad RD(b_1) = stamp \rightarrow eco.$$

Consequently, $NMAS \models counts\text{-}as^2_{\mathbf{n}}(catalytic, eco|\top)$. This formalizes that the normative system wants that if a car is catalytic, then it is considered as an ecological vehicle and the normative believes that from system putting a stamp on a catalytic car licence follows the fact that the car is catalytic. In situation $S = \{catalytic\}$, given $B_{\mathbf{n}}$ we have that the consequences of the constitutive norms are $out(B_{\mathbf{n}}, S, \geqslant_{\mathbf{n}}) = \emptyset$ and thus the goal d_3 remains unsatisfied, while in situation $S' = \{catalytic, stamp\}$ they are $out(B_{\mathbf{n}}, S', \geqslant_{\mathbf{n}}) = \{eco\}$ and the goal d_3 is satisfied.

This level of abstraction supposes that the normative system is an agent acting in the world. For example, a specialized agent introduced by the designer of the infrastructure. This abstraction can be detailed by introducing agents acting on behalf of the normative system: the normative system wants that an agent a makes the institutional fact y true if x holds in context C and believes that the effect of action z of agent a is the institutional fact y.

4.6. Fourth level of abstraction

Before introducing the next concrete level of abstraction in obligations we discuss the fourth level of constitutive norms which is based on the notion of delegation of power, since it is used in obligations.

Definition 12 *(Counts-as relation (level 4) and delegation of power).* Let *NMAS* be a normative multiagent system $\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E, \mathbf{n}, I, V, GD\rangle$.

$a \in A$ is an agent, $z \in X_a$ an action of agent a, $x \in Lit(X)$ is a literal built out of a variable, $y \in Lit(I)$ a literal built out of an institutional fact, $C \subseteq Lit(X)$ the context. Agent a has been delegated the power to consider x in context C as the institutional fact y, $NMAS \models delegated^4_{\mathbf{n}}(a, z, x, y|C)$, iff:

1. $C \cup \{x\} \rightarrow y \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(GD(a), \geqslant_{\mathbf{n}})$: it is a desire of the normative system and a goal distributed to agent a that in context C the fact x is considered as the institutional fact y.
2. $\exists z \in X_a$ such that $C \cup \{z\} \rightarrow y \in out(B_{\mathbf{n}}, \geqslant_{\mathbf{n}})$: there exists an action z of agent a such that if it decides z then the normative system believes that the institutional fact y follows (i.e., $counts\text{-}as^2_{\mathbf{n}}(z, y|C)$ at the second level of abstraction).

If $NMAS \models delegated^4_{\mathbf{n}}(a, z, x, y|C)$, then $NMAS \models counts\text{-}as^4_{\mathbf{n}}(x, y|C)$.

Example 12. $b \in A$, $P \setminus I = \{catalytic\}$, $I = \{eco\}$, $X_a = \{drive\}$, $X_b = \{stamp\}$,

$$D_{\mathbf{n}} = GD(b) = \{d_3\}, \qquad RD(d_3) = catalytic \rightarrow eco,$$
$$B_{\mathbf{n}} = \{b_1\}, \qquad RD(b_1) = stamp \rightarrow eco.$$

Thus, $NMAS \models delegated^4_{\mathbf{n}}(b, stamp, catalytic, eco|\top)$. Note that with respect to Example 11, the goal d_3 is distributed to agent b and *stamp* is an action of agent b.

We can now define obligations where agents have been delegated the power of recognizing violations by means of actions which count as such. Differently from the obligation of level 3, clause 2 distributes a goal to agent b who is in charge of recognizing violations and whose action z is believed by the normative system $\mathbf{n}$ to be the recognition of a violation (clause 7).

Definition 13 *(Obligation (level 4)).* Let *NMAS* be a normative multiagent system

$$\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E, \mathbf{n}, I, V, GD\rangle.$$

Agent $a \in A$ is *obliged* to see to it that $x \in Lit(X_a \cup P)$ with $V(\sim x, a) \in Lit(P)$ and sanction $s \in Lit(X_c \cup P)$ in context $Y \subseteq Lit(X)$ in *NMAS*, written as $NMAS \models O^4_{a\mathbf{n}}(x, s|Y)$, if and only if $\exists b, c \in A$ and a decision variable $z \in X_b$ such that Definition 10 holds except that:

2. $Y \cup \{\sim x\} \rightarrow V(\sim x, a) \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(GD(b), \geqslant_{\mathbf{n}})$: if Y holds and $\sim x$ is true, then agent $\mathbf{n}$ has distributed to agent b the goal $V(\sim x, a)$: that it is recognized as a violation in context Y.
4. $Y \cup \{V(\sim x, a)\} \rightarrow s \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(GD(c), \geqslant_{\mathbf{n}})$: if Y holds and agent $\mathbf{n}$ decides $V(\sim x, a)$, then agent $\mathbf{n}$ has distributed the goal to sanction agent a to agent c.
7. $Y \cup \{z\} \rightarrow V(\sim x, a) \in out(B_{\mathbf{n}}, \geqslant_{\mathbf{n}})$: from action z of agent b $\mathbf{n}$ believes it follows the recognition of the violation.

From clauses 2 and 7 it follows that agent b has been delegated the power to recognize violations by means of his action z:

$$\text{if } NMAS \models O^4_{a\mathbf{n}}(x, s|Y) \text{ then } NMAS \models \exists b \in A, z \in X_b \; delegated^4_{\mathbf{n}}(b, z, \sim x, V(\sim x, a) \mid Y).$$

4.7. Fifth level of abstraction (most detailed)

Clauses 2 and 4 of the definition above are like the first clauses of obligations $O_{b\mathbf{n}}(V(\sim x, a), s' \mid Y \cup \{\sim x\})$ and $O_{c\mathbf{n}}(s, s''|\sim x \wedge V(\sim x, a) \wedge Y)$. The model can thus be extended with procedural norms directed towards agents which have to take care of the procedural aspects of law, like prosecuting violations and sanctioning violators. These additional obligations are procedural norms and provide a motivation for the agents in charge of prosecuting and sanctioning. In this way, we introduce the distinction between a defender agent b who has the duty to monitor a norm

and a defender agent c who has to enforce the norm by applying sanctions: they are subject to procedural norms of the normative system.

Definition 14 *(Obligation with procedural norms (level 5)).* Let *NMAS* be a normative multiagent system $\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E, \mathbf{n}, I, V, GD\rangle$. Agent $a \in A$ is *obliged* to see to it that $x \in Lit(X_a \cup P)$ with $V(\sim x, a) \in Lit(P)$ and sanction $s \in Lit(X_c \cup P)$, $s', s'' \in Lit(X_{\mathbf{n}} \cup P)$ in context $Y \subseteq Lit(X)$ in *NMAS*, written as $NMAS \models O^5_{a\mathbf{n}}(x, s|Y)$, if and only if $\exists b, c \in A$ and a decision variable $z \in X_b$, $1 \geqslant i \geqslant 4$, such that Definition 13 holds except that:

2. $NMAS \models O^i_{b\mathbf{n}}(V(\sim x, a), s'|\sim x \wedge Y)$: if Y holds and $\sim x$ is true, then agent b is obliged by $\mathbf{n}$ that $V(\sim x, a)$ is true: that $\sim x$ is recognized as a violation in context Y.
4. $NMAS \models O^i_{c\mathbf{n}}(s, s''|V(\sim x, a) \wedge Y)$: if Y and $\sim x$ hold and b has recognized $\sim x$ as a violation done by agent a, then agent c is obliged by agent $\mathbf{n}$ that $V(\sim x, a)$ to sanction agent a with s.

Note that the procedural norms cannot be obligations defined at level 5 to avoid an infinite recursion. The idea underlying this definition is that less complex methods of control are needed to deal with the behavior of agents playing roles in the normative system. As Firozabadi and van der Torre [40] claim, at higher levels the control routines become less risky and require less effort, there is no need of a infinite regression of authorities controlling each other.

Obligations at Items 2 and 4 imply by their definitions the following goals:

1. $Y \cup \{\sim x\} \rightarrow V(\sim x, a) \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geqslant_{\mathbf{n}})$;
2. $Y \cup \{V(\sim x, a)\} \rightarrow s \in out(D_{\mathbf{n}}, \geqslant_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geqslant_{\mathbf{n}})$.

Given that these two goals are of the normative system $\mathbf{n}$ we have that:

$$NMAS \models O^5_{a\mathbf{n}}(x, s|Y) \text{ implies } NMAS \models O^4_{a\mathbf{n}}(x, s|Y).$$

Moreover, it follows that agent b has been delegated the power to recognize violations by means of his action z:

$$NMAS \models \exists b \in A, z \in X_b \; delegated^4_{\mathbf{n}}(b, z, \sim x, V(\sim x, a) \mid Y).$$

Similar definitions can be introduced for obligations where sanctions are institutional facts which have been delegated to defender agent c.

Procedural norms enters the definition of delegation of power by obliging some agent to recognize a fact as an institutional fact by exercising the delegated power:

Definition 15 *(Delegation of power with procedural norms and counts-as (level 5)).* Let *NMAS* be a normative multiagent system $\langle A, X, B, D, G, AD, E, RD, \geqslant, \geqslant_E, \mathbf{n}, I, V, GD\rangle$.

$a \in A$ is an agent, $z \in X_a$ an action of agent a, $x \in Lit(X)$ is a literal built out of a variable, $y \in Lit(I)$ a literal built out of an institutional fact, $C \subseteq Lit(X)$ the context, sanction $s \in Lit(X_{\mathbf{n}} \cup P)$ and $1 \geqslant i \geqslant 4$. Agent a has been delegated the power to consider x in context C as the institutional fact y, $NMAS \models delegated^5_{\mathbf{n}}(a, z, x, y|C)$, iff Definition 12 holds, except that:

1. $NMAS \models O^i_{b\mathbf{n}}(y, s|C \wedge x)$: agent b is obliged by the normative system $\mathbf{n}$ that in context C the fact x is considered as the institutional fact y.

Agent a can fulfill his obligations by exercising his powers specified in clause 2.
If $NMAS \models delegated^5_{\mathbf{n}}(a, z, x, y|C)$, then $NMAS \models counts\text{-}as^5_{\mathbf{n}}(x, y|C)$.
Moreover, if $NMAS \models delegated^5_{\mathbf{n}}(a, z, x, y|C)$, then $NMAS \models delegated^4_{\mathbf{n}}(a, z, x, y|C)$.

When procedural norms and agent playing roles in the system like defender agents are introduced, games become more complex. In particular an agent subject to an obligation has to predict the behavior of the defender agent of monitoring and sanctioning him. However, he knows that the defender agent will take his decision by foreseeing the

reaction of the normative system too, since he is subject to procedural norms. In [13] we discuss such more complex games.

Procedural norms allow to realize one of the key concepts of the organization of modern societies: the separation of powers as proposed in the Montesquieu's *trias politica*: the representative, executive and judicial authorities should be kept distinct [7]. Moreover decentralizing the control of the policies regulating a society supports the view that tasks can be better performed if they are dealt with by the local level in an autonomous way. As we discussed before, this delegation allows agents to interpret norms, making the system more flexible.

5. Related work

Most work in normative multiagent systems gives an abstract description of the normative system in terms of substantive norms only, specifying which is the desired social order in terms of obligations, permissions and, sometimes, counts-as norms. At this level of abstraction, the monitoring and enforcement of norms is made automatically and implicitly at the infrastructure level. Alternatively, special agents are introduced for monitoring and enforcing norms, but they are assumed to comply with their role without further motivations [19].

Moreover, no declarative representation of the monitoring and enforcement mechanisms is done. However, when the system is examined at more detailed levels of abstraction new problems arise, like which actions are necessary to establish whether an obligation has been violated, which sanctions are to be applied and, at even more detailed levels where agents acting for the normative system are introduced, who is in charge of executing these actions and which are his motivations which lead him to perform these actions. It is an open problem, moreover, how constitutive norms behave at these more detailed levels of abstraction where agents playing roles in normative systems are considered.

The framework in this paper further refine the logical framework of our game-theoretic approach to normative multiagent systems [11,13] which explicitly takes into account the activity of agents in the definition of sanction based obligations. The basic assumptions of our model are that beliefs, goals and desires of an agent are represented by conditional rules, and that, when an agent takes a decision, it recursively models the other agents interfering with it in order to predict their reaction to his decision as in a game. Most importantly, the normative system itself can be conceptualized as a socially constructed agent with whom it is possible to play games to understand what will be its reaction to the agent's decision. The actions of the normative system can be seen as abstractions of the actions of the agents to whom some of the powers of the system have been delegated.

The abstraction hierarchy explains also how the normative systems as autonomous agent metaphor introduced in our earlier work can be explained. In the model presented in [11], regulative norms are represented by the goals of the normative system and constitutive norms as its beliefs. However, when actions of the normative system and agents playing roles in it are considered new facts must be added to norms. For example, the goals of recognizing violations and applying sanctions must be added to obligations [11]. The abstraction hierarchy explains this metaphor, since at level 5 we model all details of a normative multiagent system. The abstraction at which a normative system acts as an agent is given at level 3.

Aldewereld [1] introduces an abstraction hierarchy for norm design, that is, to refine abstract norms such that they can be implemented. His approach is based on so-called "root enforcers", which we believe are not necessary [8]. It is not clear how we can map our five layers to his hierarchy. Moreover, Aldewereld does not formally define procedural norms, he only suggests that they are a kind of nested norms. Given the problems with nested norms in the context of von Wright's so-called "transmission of will", it is not clear whether nested deontic conditionals defined in the usual way are sufficient, see [13].

The levels of abstraction allow a comparison with existing models of norms in deontic logic and normative systems. Most works in normative multiagent systems give an abstract description of the normative system in terms of substantive norms only, specifying which is the desired social order in terms of obligations, permissions and, sometimes, counts-as norms, and can therefore be compared with level 1. We adopt level 2 and 3 for regulative norms in [11,13].

The notion of empowerment in normative multiagent systems is widely discussed, but it has not been related yet with the notion of goal delegation.

Pacheco and Santos [37], for example, discuss the delegation of obligations among roles. In particular, they argue that when an obligation is delegated, a corresponding permissions must be delegated too. This rationality constraint inside an institution parallels our notion of delegation of power: when the goal to make true an institutional fact is

delegated, the agent must be empowered to do so too. Moreover, in our model we can add to the notion of delegation of power also the permission for the delegated agent to perform the action which counts as the delegated institutional fact. This can be done using the definition of permission given in [13].

Pacheco and Santos consider the delegation process among roles rather than among agents. This feature can be added to our model too, using the framework for roles we discuss in [16]. Note that our model of roles describes roles by means of beliefs and goals; it is, thus, compatible with the distribution of goals to agents described by clause 2 of Definition 13.

Gelati et al. [23] combine obligations and power to define the notion of mandate in contracts: "a mandate is a proclamation intended to create the obligation of exercising a declarative power". However, they do not apply their analysis to the definition of constitutive rules but to the normative positions among agents.

Comparison with other models of counts-as is discussed in [10] and [12].

6. Further research

Some points for further research follow from our discussion on related work. For example, the use of our model for norm design could be studied, taking the work of Aldewereld as a starting point [1].

Another issue is the relation between regulative rules and delegation of power. The challenge here is to define how it is possible to create global policies [13] obliging or permitting other agents to delegate their power. Since we propose to model delegation of control by means of obligations concerning what is obligatory and what must be sanctioned, our framework can be extended with meta-policies. We can extend this framework for representing obligations by the central authority that local authorities permit of forbid access as well as permissions to forbid or permit access.

Other subjects of research can be explored when we combine the refined logical framework in this paper with other aspects of our normative multiagent systems, such as our model of roles. This allows to structure organizations in sub-organizations and roles to make a multiagent system modular and thus manage its complexity, is described in [16].

We can consider abstraction also in the underlying input/output logic. Abstraction in the input/output logic framework has been left for lions or input/output networks. In such networks each black box corresponding to an input/output logic is associated with a component in an architecture. A discussion can be found in [12].

Finally, further work is to study the use of the new refinements in applications. For example, this approach allows facing the problem of controlling distributed systems, such as virtual communities, by delegating to defender agents the task of monitoring and sanctioning violations. However, these agents are not assumed to be fully cooperative so that they are kept under the control of a judicial authority. Since, as Firozabadi and van der Torre [40] claim, at higher levels the control routines become less risky and require less effort, there is no need of a infinite regression of authorities controlling each other. As Firozabadi and Sergot [39] discuss, centralized control is not feasible in virtual communities where each participant is both a resource consumer and a resource provider. In fact, there is no authority which is in control of all the resources. Rather the central authority can only issue *meta-policies* [42] concerning the policies regulating the access to the single resources: for example, the central authority can oblige local authorities to grant access to their resources to authorized users, who are thus *entitled* to use the resources.

The model introduced in this paper also can be used for an analysis whether constitutive norms such as "cars with a catalytic converter count as ecological vehicles" can be 'violated'. The fact that obligations can be violated has been driving much research in deontic logic due to the notorious contrary-to-duty paradoxes, as well as work in normative multiagent systems where norms are considered as soft constraints rather than hard constraints. Usually it is assumed that constitutive norms cannot be violated, since they operate as a kind of definitions of the normative system. However, at a more detailed analysis proposed in this paper, it may be argued as well that the constitutive norm is 'violated' when the agent who is delegated the power to count catalytic converter as ecological vehicles does not do so. We believe that the notion of 'violation' of constitutive norms can drive further understanding and development of constitutive norms in particular and normative multiagent systems in general, and one of the results of this paper is that it is made precise in what sense and at which level of abstraction we can say that constitutive norms are 'violated'. Despite the fact that in some sense and at some level of abstraction both regulative and constitutive norms can be violated, there are still various distinctions between them, made precise by our formal analysis.

At the lower levels it becomes possible to answer the question not only of whether constitutive norms can be violated like it happens for regulative ones, but also whether constitutive norms can be interpreted. Interpreting an

obligation means that the agent in charge for monitoring violations or applying sanctions decides that in a given situation it is not worthwhile to apply the norms. This is not necessarily an abuse of his position, since norms cannot foresee all possible cases where they can applied, nor all the possible exceptions. In the same way, it is possible to interpret also constitutive rules, since the institutional facts follow from actions delegated to some agent. This agent can decide to execute the action in the required context or not. For example, a constitutive norm applying to signatures can be extended to new form of signatures, like digital ones. Conversely, a signature by a fifteen years person may not be considered as a case where the same rule can be applied, even if the law does no say explicitly anything about this case. The possibility to interpret constitutive norms, however, gives rise to new possibilities that constitutive norms are violated or abused. A constitutive norm can be violated in the sense that the normative system or the agent who is delegated the goal to recognize the institutional fact and empowered to do so fails to achieve the delegated goal. In our running example the office could fail or refuse to recognize the car as an ecological vehicle. The reason can be the inability to perform the necessary actions, laziness, bribing, etc., like it happens for regulative norms. Moreover, constitutive rules can be abused, in the sense that the delegated agent can exercise its power without being delegated to do so in the given circumstances. This possibility assumes that the institutional power can be exercised beyond the conditions under which it has been delegated the goal to exercise it.

A further issue of future research is how to define norms at different level of abstraction, rather then the normative system itself. Abstract norms can be then refined in specific institutions as discussed by [26].

7. Summary

In this paper we refine the logical framework of our game-theoretic approach to normative multiagent systems. Starting from our previous model of regulative norms [11,14], which already considers some procedural aspects, we study how also counts-as constitutive norms are associated with procedural norms. In this context we introduce the notion of delegation of power. Counts-as relations do not always provide directly an abstract recognition of brute facts in terms of institutional facts. We argue that in many cases the inference that concludes institutional facts from brute facts is the result of actions of agents acting on behalf of the normative systems and who are in charge of recognizing which institutional facts follow from brute facts. These agents make the institutional facts "official", i.e., "legally binding". The delegation of this responsibility opens space for interpretation of constitutive norms, thus making the system more flexible. Delegation of power is composed of a direct counts-as relation specifying that the effect of an action of an agent is an institutional fact and by a goal of the normative system that the fact is considered as an institutional fact. Constitutive norms are used not only to define intermediate concepts at each level, but they also enter in the definition of sanction based obligations, to define what counts as the recognition of a violation at level 4, and regulative norms are used to see to motivate agents to apply constitutive norms at level 5. The notion of delegation of power is necessary to achieve mutual empowerment, in the sense that the agents playing a role in a normative system empower this system, and the system delegates again this power to the agents. We show how counts-as relations in some cases depend on the action of agents which are in charge of recognizing facts as institutional facts. Moreover, we show that these agents are motivated to do so by a goal delegated to them by the normative system. We add procedural norms to both delegation of power and obligations. In the resulting logical framework, we can consider the optimal decision of an agent at the following five levels of abstraction (for various applications of normative multiagent systems):

1. The optimal decision of an agent follows from a simple decision problem, because it abstract from the fact that violations, sanctions and institutional facts are the result of the action that some agent decides to perform. The recognition of the violation and the sanction logically follow from the violation. The responsibility of monitoring, sanctioning and making counts-as inferences is completely delegated to the infrastructure.
2. The optimal decision takes into account the beliefs of the normative system, because the normative system is in some way personified and is attributed mental attitudes. The recognition of violations and sanctions are wanted by normative system itself. If obligations are goals of the normative system, the institutional facts follow from the beliefs of the normative system: they are still logical consequences of facts, but in the beliefs of the normative system.
3. The optimal decision can only be calculated by taking the reaction of the normative system into account, because institutional facts follow from actions of the normative system, and the recognition of violations and sanctions

are considered as the actions of the normative system itself. The normative system can decide to enforce norms or not in specific situations, thus allowing interpretation of them.

4. The optimal decision takes into account the actions of the agents acting on behalf of the normative system. Some agents are delegated the goal and power to sanction violations, to recognize violations, and to recognize some facts as institutional facts.
5. The optimal decision takes into account the actions of other agents acting on behalf of the normative system, and thereafter again the normative system, because procedural norms are used to motivate the agents playing roles in the system. These norms oblige agents playing roles to achieve the goals of the normative system. Some agents are obliged to achieve the delegated goal to sanction violations, the goal and power of recognizing violations, and to achieve the goal to recognize some facts as institutional facts by exercising the delegated powers. In the Italian law, for example, it is obligatory for an attorney to start a prosecution process when he comes to know about a crime (art. 326 of *Codice di procedura penale*).

References

[1] H. Aldewereld, Autonomy vs. conformity: An institutional perspective on norms and protocols, PhD thesis, Utrecht University, 2007.
[2] A. Anderson, A reduction of deontic logic to alethic modal logic, Mind 67 (1958) 100–103.
[3] L. Aqvist, Deontic logic, in: D. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logic: vol. II: Extensions of Classical Logic, Reidel, Dordrecht, 1984, pp. 605–714.
[4] A. Artikis, M. Sergot, J. Pitt, An executable specification of an argumentation protocol, in: Procs. of 9th International Conference on Artificial Intelligence and Law, ICAIL 2003, ACM Press, New York, 2003.
[5] A. Bochman, Explanatory Nonmonotonic Reasoning, World Scientific Publishing, London, 2005.
[6] G. Boella, L. van der Torre, Norm governed multiagent systems: The delegation of control to autonomous agents, in: Procs. of the 2003 IEEE/WIC International Conference on Intelligent Agent Technology (IAT'03), IEEE, 2003.
[7] G. Boella, L. van der Torre, Game specification in normative multiagent system: The trias politica, in: Procs. of IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'04), IEEE, 2004.
[8] G. Boella, L. van der Torre, Enforceable social laws, in: Procs. of 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05), ACM, New York, 2005.
[9] G. Boella, L. van der Torre, From the theory of mind to the construction of social reality, in: Procs. of the 27th Annual Conference of the Cognitive Science Society (CogSci'05), Lawrence Erlbaum, Mahwah, 2005.
[10] G. Boella, L. van der Torre, Constitutive norms in the design of normative multiagent systems, in: Computational Logic in Multi-Agent Systems, 6th International Workshop (CLIMA VI), in: LNCS, vol. 3900, Springer, Berlin, 2006.
[11] G. Boella, L. van der Torre, A game theoretic approach to contracts in multiagent systems, IEEE Transactions on Systems, Man and Cybernetics—Part C: Applications and Reviews 36 (1) (2006) 68–79.
[12] G. Boella, L. van der Torre, A logical architecture of a normative system, in: Deontic Logic and Artificial Normative Systems, 8th International Workshop on Deontic Logic in Computer Science (ΔEON'06), in: LNCS, vol. 4048, Springer, Berlin, 2006.
[13] G. Boella, L. van der Torre, Security policies for sharing knowledge in virtual communities, IEEE Transactions on Systems, Man and Cybernetics—Part A: Systems and Humans 36 (3) (2006) 439–450.
[14] G. Boella, L. van der Torre, Institutions with a hierarchy of authorities in distributed dynamic environments, Artificial Intelligence and Law Journal (AILaw).
[15] G. Boella, L. van der Torre, Norm negotiation in multiagent systems, International Journal of Cooperative Information Systems (IJCIS) 16 (2) (2007) 97–122 (Special Issue: Emergent Agent Societies).
[16] G. Boella, L. van der Torre, The ontological properties of social roles in multi-agent systems: Definitional dependence, powers and roles playing roles, Artificial Intelligence and Law Journal (AILaw).
[17] G. Boella, L. van der Torre, H. Verhagen, Introduction to normative multiagent systems, Computation and Mathematical Organizational Theory 12 (2–3) (2006) 71–79 (Special Issue on Normative Multiagent Systems).
[18] E. Bottazzi, R. Ferrario, A path to an ontology of organizations, in: Procs. of EDOC Int. Workshop on Vocabularies, Ontologies and Rules for The Enterprise (VORTE 2005), 2005.
[19] R. Brafman, M. Tennenholtz, On partially controlled multi-agent systems, Journal of Artificial Intelligence Research (JAIR) 4 (1996) 477–507.
[20] J. Broersen, M. Dastani, J. Hulstijn, L. van der Torre, Goal generation in the BOID architecture, Cognitive Science Quarterly 2 (3–4) (2002) 428–447.
[21] C. Castelfranchi, Modeling social action for AI agents, Artificial Intelligence 103 (1–2) (1998) 157–182.
[22] C. Castelfranchi, The micro-macro constitution of power, Protosociology 18 (2003) 208–269.
[23] J. Gelati, A. Rotolo, G. Sartor, G. Governatori, Normative autonomy and normative co-ordination: Declarative power, representation, and mandate, Artificial Intelligence and Law 12 (1–2) (2004) 53–81.
[24] M. Georgeff, F.F. Ingrand, Decision-making in an embedded reasoning system, in: Procs. of 11th International Joint Conference on Artificial Intelligence (IJCAI'89), Morgan Kaufmann, San Mateo, CA, 1989.
[25] P.J. Gmytrasiewicz, E.H. Durfee, Formalization of recursive modeling, in: Procs. of the 1st International Conference on Multiagent Systems (ICMAS'95), AAAI/MIT Press, Cambridge, 1995.

[26] D. Grossi, H. Aldewereld, J. Vazquez-Salceda, F. Dignum, Ontological aspects of the implementation of norms in agent-based electronic institutions, Computational & Mathematical Organization Theory 12 (2–3) (2006) 251–275.
[27] D. Grossi, F. Dignum, J. Meyer, Contextual terminologies, in: Computational Logic in Multi-Agent Systems, 6th International Workshop (CLIMA VI), in: LNCS, vol. 3900, Springer, Berlin, 2006.
[28] D. Grossi, J.-J. Meyer, F. Dignum, Counts-as: Classification or constitution? An answer using modal logic, in: Deontic Logic and Artificial Normative Systems, 8th International Workshop on Deontic Logic in Computer Science (ΔEON'06), in: LNCS, vol. 4048, Springer, Berlin, 2006.
[29] H. Hart, The Concept of Law, Clarendon Press, Oxford, 1961.
[30] A. Jones, J. Carmo, Deontic logic and contrary-to-duties, in: D. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logic, vol. 3, Kluwer, Dordrecht, 2001, pp. 203–279.
[31] A. Jones, M. Sergot, A formal characterisation of institutionalised power, Journal of IGPL 3 (1996) 427–443.
[32] J. Lang, L. van der Torre, E. Weydert, Utilitarian desires, Autonomous Agents and Multiagent Systems 5 (3) (2002) 329–363.
[33] D. Makinson, L. van der Torre, Input–output logics, Journal of Philosophical Logic 29 (4) (2000) 383–408.
[34] D. Makinson, L. van der Torre, Constraints for input–output logics, Journal of Philosophical Logic 30 (2) (2001) 155–185.
[35] Merriam-Webster, Dictionary of Law, Merriam-Webster, 1996.
[36] J.J.C. Meyer, A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic, Notre Dame Journal of Formal Logic 29 (1) (1988) 109–136.
[37] O. Pacheco, F. Santos, Delegation in a role-based organization, in: Procs. of ΔEON'04, Springer, Berlin, 2004.
[38] A. Ross, Tû-tû, Harvard Law Review 70 (5) (1957) 812–825.
[39] B. Sadighi Firozabadi, M. Sergot, Contractual access control, in: Security Protocols, 10th International Workshop, in: LNCS, vol. 2845, Springer, Berlin, 2004.
[40] B. Sadighi Firozabadi, L. van der Torre, Formal models of control systems, in: Procs. of 13th European Conference on Artificial Intelligence (ECAI'98), John Wiley and Sons, Chichester, 1998.
[41] J. Searle, The Construction of Social Reality, The Free Press, New York, 1995.
[42] M.S. Sloman, Policy driven management of distributed systems, Journal of Network and Systems Management 2 (4) (1994) 333–360.
[43] L. van der Torre, Y. Tan, Contrary-to-duty reasoning with preference-based dyadic obligations, Annals of Mathematics and Artificial Intelligence 27 (1–4) (1999) 49–78.
[44] G.H. von Wright, Deontic logic, Mind 60 (1950) 1–15.