Constitutive Norms in the Design of Normative Multiagent Systems

Guido Boella[1] and Leendert van der Torre[2]

[1]Dipartimento di Informatica - Università di Torino - Italy. email: guido@di.unito.it
[2]University of Luxembourg. e-mail: leendert@vandertorre.com

Abstract. In this paper, we consider the design of normative multiagent systems composed of both constitutive and regulative norms. We analyze the properties of constitutive norms, in particular their lack of reflexivity, and the trade-off between constitutive and regulative norms in the design of normative systems. As methodology we use the metaphor of describing social entities as agents and of attributing them mental attitudes. In this agent metaphor, regulative norms expressing obligations and permissions are modelled as goals of social entities, and constitutive norms expressing "counts-as" relations are their beliefs.

1 Introduction

Legal systems are often modelled using regulative norms, like obligations, prohibitions, and permissions [1]. However, a large part of the legal code does not contain obligations, prohibitions and permissions, but definitions for classifying the commonsense world under legal categories, like contract, money, property, marriage. Regulative norms can refer to this legal classification of reality.

Consider the consequences for the design of legal systems. For example, in [2] we address the issue of designing obligations to achieve the objectives of the legal system. However, the problem has not been studied of how to design legal systems composed of both constitutive and regulative norms. For modelling constitutive norms, specialized formalisms for counts-as conditionals have been introduced [3–5], but it remains unclear how to relate them to regulative norms. In contrast, as Artosi *et al.* [3] argue, for constitutive norms to be norms it is necessary that "their conditional nature exhibits some basic properties enjoyed by the usual normative links". Thus constitutive and regulative norms should be more strictly related.

Obligations, prohibitions and permissions have a conditional nature. Their conditions could directly refer to entities and facts of the commonsense world, but they can rather refer to a legal and more abstract classification of the world, making them more independent from the commonsense view. E.g., they refer to money instead of paper sheets, to properties instead of houses and fields. This more natural and economical way to model the relation between commonsense reality and legal reality uses "counts-as" conditionals, and allows regulative norms to refer to the legal classification of reality. In this way, e.g., it is not necessary that each regulative norm refers to all the conditions involved in the classification of paper as money or of houses and fields as properties. Moreover, it is not necessary that regulative norms manage the exceptions in

the classification, e.g., that a fake bill is not money or that some field is not considered as a property. Finally, by referring to the legal classification of reality only, regulative norms are not sensitive anymore to changes in the classification: a new bill can be introduced without changing the regulative norms concerning money, or a new form of property or a new kind of marriage can be introduced without changing the relevant norms.

However, the trade-off and equivalences between systems made purely of regulative norms and those including also constitutive norms cannot be easily captured by specialized formalisms. They either consider only regulative norms, such as deontic logic, or only constitutive norms, such as logics of counts-as conditionals, or, finally, with formalisms using very different formalizations for modelling the two kinds of norms. This is a problem for the design of normative systems.

In [6], to model social reality, we have introduced constitutive norms in our normative multiagent systems. In this paper we use normative multiagent systems to model the design of legal systems. In particular, the research questions of this paper are: What properties have constitutive norms? In [6] we use rules satisfying the identity property, thus making the "counts-as" relation reflexive. This is a undesired property if constitutive norms provide a classification of reality in term of legal categories. In this paper we remedy this by modelling "counts-as" as input/output conditionals. This is an alternative solution with respect to the one proposed by Artosi *et al.* [3]. Secondly, how can regulative and constitutive norms be traded-off against each other in the design of legal systems? If we replace constitutive norms in a legal system with regulative ones, then we lose the abstraction provided by legal classification.

The main advantage of our approach in comparison with other accounts, is that we combine constitutive and regulative norms in a single conceptual model. As methodology we use our model of normative multiagent systems introduced in AI and agent theory to model social reality and agent organizations [7, 8]. 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 [9] the other agents interfering with it in order to predict their reaction to its decision as in a game. Most importantly, the normative system itself can be conceptualized as an agent with whom it is possible to play games to understand what will be its reaction to the agent's decision: to consider its behavior as a violation and to sanction it. In the model presented in [6], regulative norms are represented by the goals of the normative system and constitutive norms as its beliefs. In this paper we discuss the properties of counts-as relations relating them to the properties of beliefs and how trade-off problem between constitutive and regulative norms can be handled by as the trade-off between beliefs and goals of the normative system. The cognitive motivations of the agent metaphor underlying our framework are discussed in [10].

The paper is organized as follows. In Section 2 we describe the agent metaphor. In Section 3 we introduce a logic which does not satisfy identity. In Section 4 we discuss the relation between constitutive and regulative norms. In Section 5 we introduce a formal model where we discuss the properties of constitutive norms and in Section 6 the trade-off with regulative ones. Comparison with related work and conclusion end the paper.

2 Attributing mental attitudes

We start with a well known definition:“ *Normative systems* are sets of agents (human or artificial) whose interactions can fruitfully be regarded as norm-governed; the norms prescribe how the agents ideally should and should not behave [...]. Importantly, the norms allow for the possibility that actual behaviour may at times deviate from the ideal, i.e. that violations of obligations, or of agents rights, may occur” [1].

This definition of Carmo and Jones does not seem to require that the normative system is autonomous, or that its behavior is driven by beliefs and desires.

In [6] we use the agent metaphor which attributes mental attitudes to normative systems in order to explain normative reasoning in autonomous agents. The normative system is considered as an agent with whom the bearer of the norms plays a game. Henceforth, we can call it the normative agent.

Our motivation for using the agent metaphor is inspired by the interpretation of normative *multiagent* systems as dynamic social orders. According to Castelfranchi [11], a social order is a pattern of interactions among interfering agents “such that it allows the satisfaction of the interests of some agent”. These interests can be a delegated goal, a value that is good for everybody or for most of the members; for example, the interest may be to avoid accidents. We say that agents attribute the mental attitude ‘goal’ to the normative system, because all or some of the agents have socially delegated goals to the normative system; these goals are the content of the obligations regulating it.

Moreover, social order requires *social control*, “an incessant local (micro) activity of its units” [11], aimed at restoring the regularities prescribed by norms. Thus, the agents attribute to the normative system, besides goals, also the ability to autonomously enforce the conformity of the agents to the norms, because a dynamic social order requires a continuous activity for ensuring that the normative system’s goals are achieved. To achieve the normative goal the normative system forms the subgoals to consider as a violation the behavior not conform to it and to sanction violations. Norms, however, do not aim only at regulating behavior.

Searle argues that there are two types of norms: “Some rules regulate antecedently existing forms of behaviour. For example, the rules of polite table behaviour regulate eating, but eating exists independently of these rules. Some rules, on the other hand, do not merely regulate an antecedently existing activity called playing chess; they, as it were, create the possibility of or define that activity. The activity of playing chess is constituted by action in accordance with these rules. Chess has no existence apart from these rules. The institutions of marriage, money, and promising are like the institutions of baseball and chess in that they are systems of such constitutive rules or conventions” ([12], p. 131).

According to Searle, institutional facts like marriage, money and private property emerge from an independent ontology of “brute” natural facts through constitutive norms of the form “such and such an X counts as Y in context C” where X is any object satisfying certain conditions and Y is a label that qualifies X as being something of an entirely new sort. Examples of constitutive norms are “X counts as a presiding official in a wedding ceremony”, “this bit of paper counts as a five euro bill” and “this piece of land counts as somebody’s private property”.

In our model, we define constitutive norms in terms of the normative system's belief rules and the institutional facts as the consequences of these beliefs rules.

The propositions describing the world are distinguished in two categories: first, what Searle calls "brute facts": natural facts and events produced by the actions of the agents. Second, "institutional facts": a legal classification of brute facts; they belong only to the beliefs of the normative system and have no direct counterpart in the world. Belief rules connect beliefs representing the state of the world to other beliefs which are their consequences. They have a conditional character and are represented in the same rule based formalism as goals and desires. In the case of the normative system the belief rules have as consequences not other beliefs about brute facts in the world (e.g., "if a glass drops, it breaks"), but new legal, institutional facts whose existence is related only to the normative system. These belief rules, moreover, can connect also institutional facts to other institutional facts.

This type of belief rules expresses the *counts-as* relations which are at the basis of constitutive norms. It is important that belief rules have a conditional character, since they must reflect the conditional nature of the counts-as relation as proposed by Searle: "such and such an X counts as Y in context C".

A fact p counts as an institutional fact q in context C for normative system $\mathbf{n}$ $counts\text{-}as_{\mathbf{n}}(p, q \mid C)$, iff agent $\mathbf{n}$ believes that $p \wedge C$ has q as a consequence.

The agent metaphor attributing mental attitudes to normative systems allows to understand how humans can conceive social reality by resorting to a better known domain. In [10], we discuss the cognitive basis of our model. In this way we are able to ground the ontology of social reality into a domain which can be modelled with the existing formal instruments. Most approaches, in contrast see social entities as black boxes, of which they describe the properties from an external point of view. In our model, instead, we explain the properties of normative systems as stemming from its conceptualization as an agent.

Mental attitudes of agents, however, have usually a private character: it is not possible to know which are the real goals and beliefs of an agent apart from inferring them from its behavior. In contrast, norms have a public character, otherwise it would not be possible to achieve a social order. When we map norms into beliefs and goals of the normative agent, we do not mean that they get a private character. The normative agent is only a socially constructed agent which exists only due to the collective acceptance by all the agents of the normative multiagent system.

Another advantage of considering normative systems as agents is that agents can play games with the normative system to understand whether they will be sanctioned.

The attribution of mental entities to normative systems is a methodology which can be grounded in different formal models, among which modal logic [13]. However, mental attitudes, as well as norms, are traditionally considered as conditional attitudes, thus we resort to a specialized logic which has been developed for this purpose: the Input/output logic.

We extend this approach advocated in [6] in two ways. First we give a logical analysis of counts-as, and we argue that it requires an identity free logic. Second we discuss the trade-off between the two kinds of norms.

3 Input/output logic

A disadvantage of the approach in [6] is that given the reflexivity of counts-as we have that "A counts as A", which is in contrast with our intuition and with other approaches (but see Section 7 for a discussion). In particular, since the counts-as relation classifies brute facts in legal categories, a brute fact A cannot be also a legal category: they are ontologically heterogeneous concepts, thus we keep them separate for the purpose of legal classification. We therefore want to use an identity free logic, for which we take a simplified version of the input/output logics introduced in [14, 15]. In this section we explain how it works. A rule set is a set of ordered pairs $P \rightarrow q$, where P is a set of propositional variables and q a propositional variable. 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. Makinson and van der Torre write (P, q) to distinguish input/output rules from conditionals defined in other logics, to emphasize the property that input/output logic does not necessarily obey the identity rule. In this paper we do not follow this convention.

In this paper, to keep the formal exposition simple, input and output are respectively a set of literals and a literal. In input/output logics, the input and output can be arbitrary propositional formulas, not just sets of literals and literal as we do here and additional rules for conjunction of outputs and for weakening outputs are added.

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 $\{l_1, \ldots, l_n\} \rightarrow l$. We also write $l_1 \wedge \ldots \wedge l_n \rightarrow l$ and when $n = 0$ we write $\top \rightarrow l$. 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 $MD : Q \rightarrow Rul(X)$ is a total function from the pointers to the set of rules built from X.

Let $S = MD(Q)$ be a set of rules $\{P_1 \rightarrow q_1, \ldots, P_n \rightarrow q_n\}$, 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:

$$\frac{p \rightarrow r}{p \wedge q \rightarrow r} SI \qquad \frac{p \wedge q \rightarrow r, p \wedge \neg q \rightarrow r}{p \rightarrow r} OR \qquad \frac{p \rightarrow q, p \wedge q \rightarrow r}{p \rightarrow r} CT \qquad \frac{}{p \rightarrow p} Id$$

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 (simple-minded 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 $L \rightarrow l \in out(Q)$, and $l \in out^+(Q, L)$ iff $L \rightarrow l \in out^+(Q)$.

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

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

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 [15]. 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.

In input/output logics under constraints, a set of mental attitudes and an input does not have a set of propositions as output, but a set of set of propositions. We can infer a set of propositions by for example taking the join (credulous) or meet (sceptical), or something more complicated. Besides, we can adopt an output constraint (the output has to be consistent) or an input/output constraint (the output has to be consistent with the input). In this paper we only consider the input/output constraints. The following definition is inspired by [16] where we extend constraints with priorities:

Definition 2 (Constraints).
Let $\geq$: $2^Q \times 2^Q$ be a transitive and reflexive partial relation on the powerset of the pointers to rules containing at least the subset relation. Moreover, let out be an input/output logic. We define:

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

In case of contrary to duty obligations, the input represents something which is inalterably true, and an agent has to ask himself which rules (output) this input gives rise to: even if the input should have not come true, an agent has to "make the best out of the sad circumstances" [17].

Example 2. Let $MD(\{a, b, c\}) = \{a = (\top \rightarrow m), b = (p \rightarrow n), c = (o \rightarrow \neg m)\}$, $\{b, c\} > \{a, b\} > \{a, c\}$, where by $A > B$ we mean as usual $A \geq B$ and $B \ngeq A$.
$maxfamily(Q, \{o\}) = \{\{a, b\}, \{b, c\}\}$,
$preffamily(Q, \{o\}, \geq) = \{\{b, c\}\}$,
$outfamily(Q, \{o\}, \geq) = \{\{\neg m\}\}$

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): $\neg m$.

Due to space limitations we have to be brief on details with respect to input/output logics, see [14, 15] for the semantics of input/output logics, further details on its proof theory, its possible translation to modal logic, alternative constraints, and examples.

4 Constitutive norms vs regulative norms

Why are constitutive norms needed in a normative system? In [6], we argue that, first, regulative norms are not categorical, but conditional: they specify all their applicability conditions. In case of complex and rapidly evolving systems new situations arise which should be considered in the conditions of the norms. Thus, new regulative norms must be introduced or existing ones revised each time the applicability conditions must be extended to include new cases. In order to avoid changing existing norms or adding new ones, it would be more economic that regulative norms could factor out particular cases and refer, instead, to more abstract concepts only. Hence, the normative system should include some mechanism to introduce new institutional categories of abstract entities for classifying possible states of affairs. Norms could refer to this institutional classification of reality rather than to the commonsense classification: changes to the conditions of the norms would be reduced to changes to the institutional classification of reality. Second, the dynamics of the social order which the normative system aims to achieve is due to the evolution of the normative system over time, which introduces new norms, abrogates outdated ones, and, as just noticed, changes its institutional classification of reality. So the normative system must specify how the normative system itself can be changed by introducing new regulative norms and new institutional categories, and specify by whom the changes can be done. This second aspect has been addressed in [7].

In this paper we discuss how constitutive norms, even if they can be replaced by regulative norms, allow to create a level of abstraction to which regulative norms can refer to, making to less sensitive to the changes in the legal system. The cons of introducing constitutive norms is that new rules are necessary, so that a trade-off must be found between the need of abstraction and the complexity of the normative system.

As a running example, consider a society where the fact that a field has been fenced by an agent counts as the fact that the field is property of that agent. In our model this relation is expressed as a belief attributed to the normative system. The fence is a physical “brute” fact, while the fact that it is a property of someone is only an institutional fact attributed to the beliefs of the normative system.

Assume now that the normative system has as goals that if a field is fenced, no one enters it and that if a fenced field is entered, this action is considered as a violation and the violation is sanctioned. These goals form an obligation not to trespass a fenced field. However, the same legal system could have been designed in a different way using the constitutive norm above: a fenced field counts as property. The constitutive norm introduces the legal category of property which an obligation not to trespass a property can refer to: it is obligatory not to trespass property. The two legal systems are equivalent in the sense that in the same situation, the same violations hold; on the other hand, they are different since the latter introduce a legal classification of reality; thus, the obligation has as condition the institutional fact that the field is a property: the field being a property is an institutional fact believed by the normative system, while entering the field is a brute fact.

Analogously, in the purely regulative legal system, a permission to enter a fenced field if it is close to a river could be added. This permission is an exception to the obligation not to trespass fenced fields. In the second legal system, the same purpose can be reached by adding a constitutive norm which states that a field close to the river,

albeit fenced, is not a property. Note that this is different from saying that a field on the river is a property that can be trespassed, a fact which is expressed by a permission to enter a property close to the river.

The possibility that institutional facts appear as conditions in the goals of the normative system or as goals themselves explains the following puzzling assertion of Searle [18]: "constitutive rules constitute (and also regulate) an activity the existence of which is logically dependent on the rules" (p.34). How can constitutive rules *regulate* an activity, if this is the role played by regulative rules? E.g., Hindriks [19] argues that constitutive rules consist of also regulative ones.

In our model constitutive norms regulate a social activity since they create institutional facts that are conditions or objects of regulative norms. In our metaphorical mapping regulative norms are goals, and goals base their applicability in a certain situation on the beliefs of the agent: if the beliefs change, the goals which the agent pursues change too. Analogously, the institutional facts which are the consequences of constitutive rules determine what is obligatory, since the institutional facts determine which regulative rules are applicable. In the previous example, being a property indirectly regulates the behavior of agents, since entering a field is a violation only if it is a property; if a field is not a property, the goal of considering trespassing a violation does not apply.

Searle [18] interprets the creation institutional facts also in terms of what he calls "status functions": "the form of the assignment of the new status function can be represented by the formula 'X counts as Y in C'. This formula gives us a powerful tool for understanding the form of the creation of the institutional fact, because the form of the collective intentionality is to impose that status and its function, specified by the Y term, on some phenomenon named by the X term", p.46.

Where "the ascription of function ascribes *the use to which we intentionally put* these objects", p.20. Functions are usually defined in relation to goals. In our model, this teleological aspect of the notion of function depends on the fact that institutional facts make conditional goals relevant as they appear in the conditions of regulative norms or as goals themselves. The aim of fencing a field is to prevent trespassing: the obligation defines the function of property, since it is defined in terms of goals of the normative system. Hence, Searle's assertion that "the institutions [...] are systems of such constitutive rules" is partial: institutions are systems where constitutive (i.e., beliefs) and regulative (i.e., goals) rules interacts. In our model, they interplay in the same way as goals and beliefs do in agents.

From a knowledge representation point of view, constitutive norms behave as *data abstraction* in programming languages: types are gathered in new abstract data types; new procedures are defined on the abstract data types to manipulate them. So it is possible to change the implementation of the abstract data type without modifying the programs using those procedures. In our case, we have that regulative norms can be defined on abstract institutional facts: it is possible to change the constitutive norms defining the institutional facts without modifying the regulative norms which refer to those institutional facts. This analogy supports also our decision not to require identity as a property of counts-as. Data abstraction allows to hide the details concerning the implementation of the data type. Analogously, if the institutional facts are abstractions of the reality, they should hide the details consisting in the brute facts.

5 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 the BOID architecture, obligations are not taken as primitive concepts. Beliefs, desires and goals are represented by conditional rules rather then in a modal framework. We use in our model only goals rather than intentions since we consider only on decision step instead of having plans for the future moves.

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.* [21].

Given the same set of mental attitudes, agents reason and act differently: when facing a conflict among their motivations and beliefs, different agents prefer to fulfill different goals and desires. We express these agent characteristics by a priority relation on the mental attitudes which encode, as detailed in [20], how the agent resolves its conflicts. 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.

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

- *the agents A, propositional variables X, agent beliefs B, desires D and goals G are five finite disjoint sets.*
- *B, D, G are sets of pointers to rules. We write $M = D \cup G$ for the motivations defined as the union of the desires and goals.*
- *an agent description $AD : A \rightarrow 2^{X \cup B \cup M}$ 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 $P = X \setminus \cup_{b \in A} X_b$.*
- *a priority relation $\geq$: $A \rightarrow 2^{M \cup B} \times 2^{M \cup B}$ 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 $\geq_b$ for $\geq (b)$.*
  Since goals have priority over desires we have that given $S, S' \subseteq M$, for all $a \in A$, $S >_a S'$ if $S \setminus S' \subseteq G$ and $S' \setminus S \subseteq D$.

Example 3. $A = \{\mathbf{a}\}$, $X_{\mathbf{a}} = \{trespass\}$, $P = \{s, fenced\}$, $D_{\mathbf{a}} = \{d_1, d_2\}$, $\geq_{\mathbf{a}} = \{d_2\} \geq \{d_1\}$. There is a single agent, agent $\mathbf{a}$, who can trespass a field. Moreover, it can be sanctioned and the field can be fenced. It has two desires, one to trespass (d_1), another one not to be sanctioned (d_2). The second desire is more important.

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

Definition 4 (Multiagent system). *A multiagent system is a tuple $\langle A, X, B, D, G, AD, MD, \geq \rangle$, where $\langle A, X, B, D, G, AD, \geq \rangle$ is an agent set, and the*

mental description $MD : (B \cup M) \rightarrow Rul(X)$ *is a total function from the sets of beliefs, desires and goals to the set of rules built from* X*. For a set of mental attitudes* $S \subseteq B \cup M$*, we write* $MD(S) = \{MD(q) \mid q \in S\}$.

Example 4 (Continued). $MD(d_1) = \top \rightarrow trespass$, $MD(d_2) = \top \rightarrow \neg s$.

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. $V_{\mathbf{a}}(x)$ represents the decision of agent $\mathbf{n}$ that recognizes x as a violation by agent $\mathbf{a}$. The goal distribution $GD(\mathbf{a}) \subseteq G_{\mathbf{n}}$ represents the goals of agent $\mathbf{n}$ the agent $\mathbf{a}$ is responsible for.

Definition 5 (Normative system). *A normative multiagent system, written as* $NMAS$*, is a tuple* $\langle A, X, B, D, G, AD, MD, \geq, \mathbf{n}, I, V, GD\rangle$ *where the tuple* $\langle A, X, B, D, G, AD, MD, \geq\rangle$ *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* $V : Lit(X) \times A \rightarrow X_{\mathbf{n}} \cup P$ *is a function from the literals and the agents to the decision variables of the normative system and the parameters. We write* $V_a(x)$ *for* $V(x, a)$.
- *the goal distribution* $GD : A \rightarrow 2^{G_{\mathbf{n}}}$ *is a function from the agents to the powerset of the goals of the normative system, such that if* $L \rightarrow l \in MD(GD(\mathbf{a}))$*, then* $l \in Lit(X_{\mathbf{a}} \cup P)$.

Agent $\mathbf{n}$ is a normative system who has the goal that fenced fields are not trespassed.

Example 5 (Continued). There is agent $\mathbf{n}$, representing the normative system.
$X_{\mathbf{n}} = \{s, V_{\mathbf{a}}(trespass)\}$, $P = \{fenced\}$, $D_{\mathbf{n}} = G_{\mathbf{n}} = \{g_1\}$, $MD(g_1) = \{fenced \rightarrow \neg trespass\}$, $GD(\mathbf{a}) = \{g_1\}$.
Agent $\mathbf{n}$ can sanction agent $\mathbf{a}$, because s is no longer a parameter but a decision variable. $V_{\mathbf{a}}(trespass)$ represents the fact that the normative system considers a violation the action of $\mathbf{a}$ trespassing the field. It has the goal that fenced fields are not trespassed, and it has distributed this goal to agent $\mathbf{a}$.

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.

Regulative norms are conditional obligations with an associated sanction and conditional permissions. The definition of obligation contains several clauses. The first and central clause of our definition 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 $out(GD(\mathbf{a}), \geq_{\mathbf{n}})$.

The second and third clause can be read as "the absence of p is considered as a violation". The association of obligations with violations is inspired by Anderson's reduction of deontic logic to alethic logic [22]. 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 $\top \rightarrow V_{\mathbf{a}}(\sim x) \notin out(D_{\mathbf{n}}, \geq_{\mathbf{n}})$.

The fourth and fifth clause relate violations to sanctions. The fourth clause says that the normative system is motivated not to count behavior as a violation and apply sanctions 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. The second and fourth clauses can be considered as instrumental norms [23] contributing to the achievement of the main goal of the norm.

Definition 6 (Obligation). *Let $NMAS = \langle A, X, B, D, G, AD, MD, \geq, \mathbf{n}, I, V, GD\rangle$ be a normative multiagent system. Agent $\mathbf{a} \in A$ is obliged to see to it that $x \in Lit(X_{\mathbf{a}} \cup P)$ with sanction $s \in Lit(X_{\mathbf{n}} \cup P)$ in context $Y \subseteq Lit(X)$ in $NMAS$, written as $NMAS \models O_{\mathbf{an}}(x, s|Y)$, if and only if:*

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

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 modelled as an agent.

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.

Since conditions of obligations are sets of decision variables and parameters, institutional facts can be among them. In this way it is possible that regulative norms refer to institutional abstractions of the reality rather than to physical facts only.

Example 6 (Continued). Let: $\{g_1, g_2, g_4\} = G_{\mathbf{n}}$, $G_{\mathbf{n}} \cup \{g_3, g_5\} = D_{\mathbf{n}}$, $\{g_1\} = GD(\mathbf{a})$

$MD(g_2) = \{fenced, trespass\} \rightarrow V_{\mathbf{a}}(trespass)$ $MD(g_3) = \top \rightarrow \neg V_{\mathbf{a}}(trespass)$
$MD(g_4) = \{fenced, V_{\mathbf{a}}(trespass)\} \rightarrow s$ $MD(g_5) = fenced \rightarrow \sim s$
$NMAS \models O_{\mathbf{an}}(\neg trespass, s \mid fenced)$, since:

1. $fenced \rightarrow \neg trespass \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}}) \cap out(GD(\mathbf{a}), \geq_{\mathbf{n}})$
2. $\{fenced, trespass\} \rightarrow V_{\mathbf{a}}(trespass) \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geq_{\mathbf{n}})$
3. $\top \rightarrow \neg V_{\mathbf{a}}(trespass) \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}})$
4. $\{fenced, V_{\mathbf{a}}(trespass)\} \rightarrow s \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geq_{\mathbf{n}})$
5. $fenced \rightarrow \sim s \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}})$
6. $fenced \rightarrow \sim s \in out(D_{\mathbf{a}}, \geq_{\mathbf{a}})$

Permissions are defined as exceptions to obligations [16], and can be overridden by obligations in turn. A permission to do x is an exception to an obligation not to do x if agent $\mathbf{n}$ has the goal that x is not considered as a violation under some condition. The permission overrides the prohibition if the goal that something does not count as a violation ($Y \wedge x \rightarrow \neg V_{\mathbf{a}}(x)$) has higher priority in the ordering $\geq_{\mathbf{n}}$ on goal and desire rules with respect to the goal of a corresponding prohibition that x is considered as a violation ($Y' \wedge x \rightarrow V_{\mathbf{a}}(x)$):

Definition 7 (Permission). *Agent $\mathbf{a} \in A$ is permitted by agent $\mathbf{n}$ to see to it that $x \in Lit(X_{\mathbf{a}} \cup P)$ under condition $Y \subseteq Lit(X)$, written as $NMAS \models P_{\mathbf{an}}(x \mid Y)$, iff $Y \cup \{x\} \rightarrow \neg V_{\mathbf{a}}(x) \in out(G_{\mathbf{n}}, \geq_{\mathbf{n}})$: if Y and x then agent $\mathbf{n}$ wants that x is not considered a violation by agent $\mathbf{a}$.*

Example 7 (Continued). Let $P = \{fenced, river\}$, $\{g_6\} > \{g_2\}$,
$MD(g_6) = \{fenced, river, trespass\} \rightarrow \sim V_{\mathbf{a}}(trespass)$
Then $\{fenced, river, trespass\} \rightarrow \sim V_{\mathbf{a}}(trespass) \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geq_{\mathbf{n}})$
Hence, $NMAS \models P_{\mathbf{an}}(trespass \mid fenced \wedge river)$

Constitutive norms introduce new abstract categories of existing facts and entities, called institutional facts. We formalize the counts-as conditional as a belief rule of the normative system $\mathbf{n}$. Since the condition x of the belief rule is a variable it can be an action of an agent, a brute fact or an institutional fact. So, the counts-as relation can be iteratively applied.

Definition 8 (Counts-as relation). *Let $NMAS = \langle A, X, B, D, G, AD, MD, \geq, \mathbf{n}, I, V, GD \rangle$ be a normative multiagent system. A literal $x \in Lit(X)$ counts-as $y \in Lit(I)$ in context $C \subseteq Lit(X)$, $NMAS \models counts\text{-}as_{\mathbf{n}}(x, y|C)$, iff $C \cup \{x\} \rightarrow y \in out(B_{\mathbf{n}}, \geq_{\mathbf{n}})$: if agent $\mathbf{n}$ believes C and x then it believes y.*

Example 8. $P \setminus I = \{fenced\}$, $I = \{property\}$, $X_{\mathbf{a}} = \{trespass\}$, $B'_{\mathbf{n}} = \{b'_1\}$, $MD(b'_1) = fenced \rightarrow property$

Consequently, $NMAS \models counts\text{-}as_{\mathbf{n}}(fenced, property|\top)$. This formalizes that for the normative system a fenced field counts as the fact that the field is a property of that agent. The presence of the fence is a physical "brute" fact, while being a property is an institutional fact. In situation $S = \{fenced\}$, given $B'_{\mathbf{n}}$ we have that the consequences of the constitutive norms are $out(B'_{\mathbf{n}}, S, \geq_{\mathbf{n}}) = \{property\}$

As shown in the example, the logic of constitutive norms does not satisfy identity: $fenced$ is not a consequence, since it represents a brute fact and not an institutional fact. Constitutive norms, in contrast, provide a legal classification of reality in terms of institutional facts only.

The institutional facts can appear in the conditions of regulative norms as the following example shows.

Example 9 (Continued). A regulative norm which forbids trespassing can refer to the abstract concept of property rather than to fenced fields: $O_{\mathbf{an}}(\neg trespass, s \mid property)$.
Let: $\{g'_1, g'_2, g'_4\} = G'_{\mathbf{n}}$, $G'_{\mathbf{n}} \cup \{g'_3, g'_5\} = D'_{\mathbf{n}}$, $\{g'_1\} = GD(\mathbf{a})$
$MD(g'_1)$=$property \rightarrow \neg trespass$ $MD(g'_2)$=$\{property, trespass\} \rightarrow V_{\mathbf{a}}(trespass)$
$MD(g'_3) = \top \rightarrow \neg V_{\mathbf{a}}(trespass)$ $MD(g'_4) = \{property, V_{\mathbf{a}}(trespass)\} \rightarrow s$
$MD(g'_5) = property \rightarrow \sim s$
Then:

1. $property \rightarrow \neg trespass \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}}) \cap out(GD(\mathbf{a}), \geq_{\mathbf{n}})$
2. $\{property, trespass\} \rightarrow V_{\mathbf{a}}(trespass) \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geq_{\mathbf{n}})$
3. $\top \rightarrow \neg V_{\mathbf{a}}(trespass) \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}})$
4. $\{property, V_{\mathbf{a}}(trespass)\} \rightarrow s \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}}) \cap out(G_{\mathbf{n}}, \geq_{\mathbf{n}})$
5. $property \rightarrow \sim s \in out(D_{\mathbf{n}}, \geq_{\mathbf{n}})$
6. $property \rightarrow \sim s \in out(D_{\mathbf{a}}, \geq_{\mathbf{a}})$

As the system evolves, new cases can be added to the notion of property by means of new constitutive norms, without changing the regulative norms about property. E.g., if a field is inherited, then it is property of the heir: $inherit \rightarrow property \in MD(B_{\mathbf{n}})$.

Since counts-as rules are beliefs and the logic is non-monotonic due to the priority ordering on the beliefs, counts-as can be used to express exceptions to the classification thus mirroring the relation between obligations and permissions as exceptions [2].

6 The trade-off between constitutive and regulative norms

In this section, we extend our scenario described in Example 8-9 to design a legal system equivalent to the one of Example 6-7.

Example 10 (Continued). $B'_{\mathbf{n}} = \{b'_2\}$, $\{b'_2\} > \{b'_1\}$,
$MD(b'_2) = fenced \wedge river \rightarrow \neg property$.
$out(B'_{\mathbf{n}} = \{b'_1, b'_2\}, \geq_{\mathbf{n}}) = \{\{fenced \wedge river \rightarrow \neg property\}\}$ since
$maxfamily(B'_{\mathbf{n}}, S = \{fenced, river\}) = \{\{b'_1\}, \{b'_2\}\}$,
$preffamily(B'_{\mathbf{n}}, S = \{fenced, river\}, \geq_{\mathbf{n}}) = \{\{b'_2\}\}$,
$outfamily(B'_{\mathbf{n}}, S = \{fenced, river\}, \geq_{\mathbf{n}}) = \{\{\neg property\}\}$

Thus, $NMAS \models counts\text{-}as_{\mathbf{n}}(fenced, \neg property \mid river)$ and this belief overrides the former one behind $counts\text{-}as_{\mathbf{n}}(fenced, property \mid \top)$. This formalizes that the normative system does not consider as a property a fenced field if it is close to a river.

We show how a system containing constitutive and regulative norms like in Example 8-10 can be interchanged with an equivalent system of regulative norms only like the one of Example 6-7. By equivalence we mean that in the same state of the world the same violations hold. Since it is possible to replace constitutive norms with regulative norms only, a trade-off can be found between adding constitutive norms and achieving a sufficient level of abstraction.

Even if input/output logic is an inference system on rules we cannot directly prove the equivalence on the rules defining regulative and constitutive norms since they refer

to different sets of rules: goal rules and belief rules. We provide the equivalence in an indirect way by considering the combined output of the rules.

Given the operation out, we define a combined output relation: $output(Q, Z, S, \geq_{\mathbf{n}}) = out(Z, out(Q, S, \geq_{\mathbf{n}}) \cup S, \geq_{\mathbf{n}})$ where $Q \subseteq B_{\mathbf{n}}$, $Z \subseteq M_{\mathbf{n}}$ and $S \subset Lit(X \setminus I)$. The institutional facts are the result of the reasoning of the normative system, so they cannot be present in the initial state composed of brute facts.

Note that we reintroduce the brute facts S as the input of the output operation on the motivations Z since the output operation on beliefs does not satisfy identity. We need S since the conditions of regulative norms can refer to brute facts as well as to the institutional facts which are the consequences of the constitutive norms. In this way we distinguish between the legal classification of reality and the information concerning commonsense, among which the brute facts which are the input to constitutive norms. Even if we attribute belief rules to the normative system these must be distinguished from the belief rules of agents: these belief rules concern the relation between brute facts and constitute their commonsense view of the work. The normative system as agent, in contrast, does not contain any knowledge of this kind. The relevant commonsense inferences are performed by the real agents playing roles in the normative system.

In our examples we have:$output(B_{\mathbf{n}}, G_{\mathbf{n}}, S, \geq_{\mathbf{n}}) = output(B'_{\mathbf{n}}, G'_{\mathbf{n}}, S, \geq_{\mathbf{n}})$ for any $S \in Lit(X \setminus I)$.

Sketch of proof. We consider only the cases where the conditions of the goals and beliefs are satisfied. First, the normative system made of regulative norms only:

$output(B_{\mathbf{n}}, G_{\mathbf{n}}, S = \{fenced, trespass\}, \geq_{\mathbf{n}}) = out(G_{\mathbf{n}}, out(B_{\mathbf{n}}, S, \geq_{\mathbf{n}}) \cup S, \geq_{\mathbf{n}}) =$
$\{\neg trespass, V_{\mathbf{a}}(trespass), s\}$
from g_1, g_2, g_4, where $out(B_{\mathbf{n}}, S, \geq_{\mathbf{n}}) = \emptyset$ since $B_{\mathbf{n}} = \emptyset$.

In contrast:

$output(B_{\mathbf{n}}, G_{\mathbf{n}}, S{=}\{fenced, river, trespass\}, \geq_{\mathbf{n}}) =$
$out(G_{\mathbf{n}}, out(B_{\mathbf{n}}, S, \geq_{\mathbf{n}}) \cup S, \geq_{\mathbf{n}}) = \{\neg trespass, \neg V_{\mathbf{a}}(trespass), \sim s\}$
(from g_1, g_5, g_6) where again $out(B_{\mathbf{n}}, S, \geq_{\mathbf{n}}) = \emptyset$.

In case of the legal system of Example 8 made of both constitutive and regulative norms:

$output(B'_{\mathbf{n}}, G'_{\mathbf{n}}, S{=}\{fenced, trespass\}, \geq_{\mathbf{n}}){=}out(G'_{\mathbf{n}}, out(B'_{\mathbf{n}}, S, \geq_{\mathbf{n}}) \cup S, \geq_{\mathbf{n}}) =$
$\{\neg trespass, V_{\mathbf{a}}(trespass), s\}$
(from g'_1, g'_2, g'_4) where $out(B'_{\mathbf{n}}, S, \geq_{\mathbf{n}}) = \{property\}$ (from b'_1).

In contrast:

$output(B'_{\mathbf{n}}, G'_{\mathbf{n}}, S{=}\{fenced, river, trespass\}, \geq_{\mathbf{n}}) =$
$out(G'_{\mathbf{n}}, out(B'_{\mathbf{n}}, S, \geq_{\mathbf{n}}) \cup S, \geq_{\mathbf{n}}){=}\{\neg trespass, \neg V_{\mathbf{a}}(trespass), \sim s\}$
(from g'_1, g'_3, g'_5) where $out(B'_{\mathbf{n}}, S, \geq_{\mathbf{n}}) = \{\neg property\}$ (from b'_2).

In summary, the trade-off between constitutive and regulative rules has to take into considerations, first, how many regulative rules share the same conditions. The design of the system of norms can be simplified by introducing abstractions representing the overlapping conditions. Second how frequently the normative system is updated. In case of dynamic situations, the preferred design of the system introduces constitutive rules introducing institutional facts which are abstractions which hide the details concerning the brute facts. In this way, new cases can be dealt with without changing the regulative part of the system, but only revising what counts as an institutional fact.

7 Related work

While the formalization of regulative norms, like obligations, prohibitions and permissions, is often based in deontic logic on modal operators representing what is obligatory, forbidden or permitted, the formalization of constitutive norms is rather different. An attempt to make the notion of constitutive norm more precise is Jones and Sergot [5]'s formalization of the counts-as relation. For Jones and Sergot, the counts-as relation expresses the fact that a state of affairs or an action of an agent "is a sufficient condition to guarantee that the institution creates some (usually normative) state of affairs". As Jones and Sergot suggest, this relation can be considered as "constraints of (operative in) [an] institution", and they express these constraints as conditionals embedded in a modal operator. Jones and Sergot formalize this introducing a conditional connective $\Rightarrow_s$ to express the "counts-as" connection holding in the context of an institution s. They characterise the logic for $\Rightarrow_s$ as a classical conditional logic plus the axioms:

$((A \Rightarrow_s B) \wedge (A \Rightarrow_s C)) \supset (A \Rightarrow_s (B \wedge C))$
$((A \Rightarrow_s B) \wedge (C \Rightarrow_s B)) \supset ((A \vee C) \Rightarrow_s B)$
$((A \Rightarrow_s B) \wedge (B \Rightarrow_s C)) \supset (A \Rightarrow_s C)$

In addition, Jones and Sergot's analysis is integrated by introducing the normal KD modality D_s such that D_sA means that A is "recognised by the institution s". Accordingly, it is adopted the schema: $(A \Rightarrow_s B) \supset D_s(A \supset B)$.

The limitation of this approach, according to Gelati *et al.* [24], is that the consequences of counts-as connections follow non-defeasibly (via the closure of the logic for modality D_s under logical implication), whereas defeasibility seems a key feature of such connections. The classical example is that in an auction if a person raises one hand, this may count as making a bid. However, this does not hold if he raises his hand and scratches his own head.

Finally, the adoption of the transitivity for their logic is criticized by Artosi *et al.* [3]. Artosi *et al.* [3]'s characterisation of the counts-as adopts a different perspective. Rather than introducing a logic for the counts-as connection, and then linking it with a D_s logic, they use one conditional operator $\Rightarrow$ to express any defeasible normative connections in any institutions. They use the same D_s operator as in [5] but they apply it to the components of normative links, to relativise them to a particular institution. Any institution can only state what normative situation holds for itself, given certain conditions, but according to a general type of conditionality. On the basis of $\Rightarrow$ they define a relativised $\Rightarrow_s$ operator: $(A \Rightarrow_s B) =_{def} (A \Rightarrow D_sB) \wedge (D_sA \Rightarrow D_sB)$

The connective $\Rightarrow$ is characterised by reflexivity and cumulative transitivity, whose combination does not prevent defeasibility. The system is completed by introducing a restricted version of the detachment of the consequent. To avoid losing non-monotonicity, Artosi *et al.* [3] do not accept the strengthening of antecedent property (SI in our input/output logic), thus making their logic weaker.

In contrast, in our model we accept the strengthening of antecedent (SI) rule and the cumulative transitivity (CT). We do not accept instead identity (Id). First of all, the adoption also of Id would make the system accepting also full transitivity. Non-monotonicity is achieved via the constraint mechanism which uses also a priority ordering on the mental attitudes. Secondly, we do not accept Id because we want to keep

separate brute facts and institutional facts "whose nature - as also Artosi *et al.* [3] accept - is conceptually distinct from that of the empirical facts".

Our position is congruent also with Castelfranchi and Tummolini [25] who argue that counts-as rules regulate a cognitive activity, *viz.* the proper application of a concept:

> A constitutive rule describes, albeit very abstractly, a recognition process. [...] The application of a concept in fact can be represented in form of a rule that associates a specific set of stimuli ("something such and such") X with a linguistic label Y.

Since the stimuli and the linguistic label Y are ontologically heterogeneous, the "counts-as" relation cannot be reflexive.

Grossi and colleagues [26, 27] develop a notion of counts-as as a contextual classification in a modal logic setting, where for the classification aspect they use either description logic [26] or plain propositional logic [27]. They end up with a very strong logic for counts-as, satisfying rules not satisfied by Jones and Sergot's logic or the logic proposed in this paper, such as the identity rule (x counts-as x). They argue that the new rules are explained by their particular concept of counts-as as a contextual classification.

8 Conclusions

In this paper we discuss the design of legal systems composed of constitutive and regulative norms. We model legal systems as normative multiagent systems where the normative system is modelled as an agent using the agent metaphor: constitutive norms are defined by the beliefs of the normative system and the regulative norms by its goals. The characteristic of the counts-as relation is that it is not reflexive. The trade-off problem between constitutive and regulative norms can be handled by as the trade-off between beliefs and goals of the normative system. We show that constitutive norms, even if they can be replaced by regulative norms, allow to create a level of abstraction to which regulative norms can refer to, making it less sensitive to the changes in the legal system.

In [6] we extend this framework to model the problem of how the normative system itself specifies who can change the normative system. This specification is made by means of constitutive norms describe what facts count as the creation of new regulative and constitutive norms in the normative system. This work is at the basis of the definition of contracts we make in [7]. Future work is, for example, elaborating the notion of context to study which properties hold for it, and introducing hierarchies of normative systems composed of both constitutive norms and regulative norms, as we do for obligations and permissions in [16]. Moreover in [8] we discuss global policies about local policies in secure knowledge management. However, it has still to be studied global policies about constitutive rules.

References

1. Jones, A., Carmo, J.: Deontic logic and contrary-to-duties. In Gabbay, D., Guenthner, F., eds.: Handbook of Philosophical Logic. Kluwer (2001) 203–279
2. Boella, G., van der Torre, L.: Permissions and obligations in hierarchical normative systems. In: Procs. of ICAIL'03, New York (NJ), ACM Press (2003) 109–118
3. Artosi, A., Rotolo, A., Vida, S.: On the logical nature of count-as conditionals. In: Procs. of LEA 2004 Workshop. (2004)
4. Grossi, D., Dignum, F., Meyer, J.J.: Contextual taxonomies. In: LNCS n. 3487: Procs. of CLIMA'04 Workshop, Berlin, Springer Verlag (2004) 33–51
5. Jones, A., Sergot, M.: A formal characterisation of institutionalised power. Journal of IGPL **3** (1996) 427–443
6. Boella, G., van der Torre, L.: Regulative and constitutive norms in normative multiagent systems. In: Procs. of 10th International Conference on the Principles of Knowledge Representation and Reasoning KR'04, Menlo Park (CA), AAAI Press (2004) 255–265
7. Boella, G., van der Torre, L.: A game theoretic approach to contracts in multiagent systems. IEEE Transactions on Systems, Man and Cybernetics - Part C (2006)
8. Boella, G., van der Torre, L.: Security policies for sharing knowledge in virtual communities. IEEE Transactions on Systems, Man and Cybernetics - Part A (2006)
9. Gmytrasiewicz, P.J., Durfee, E.H.: Formalization of recursive modeling. In: Procs. of ICMAS'95, Cambridge (MA), AAAI/MIT Press (1995) 125–132
10. Boella, G., van der Torre, L.: From the theory of mind to the construction of social reality. In: Procs. of CogSci'05, Mahwah (NJ), Lawrence Erlbaum (2005) 298–303
11. Castelfranchi, C.: Engineering social order. In: LNCS n.1972: Procs. of ESAW'00, Berlin, Springer Verlag (2000) 1–18
12. Searle, J.: Speech Acts: an Essay in the Philosophy of Language. Cambridge University Press, Cambridge (UK) (1969)
13. Boella, G., van der Torre, L.: Obligations as social constructs. In: LNAI n. 2829: AI*IA 2003 - Advances in Artificial Intelligence, Berlin, Springer Verlag (2003) 27–38
14. Makinson, D., van der Torre, L.: Input-output logics. Journal of Philosophical Logic **29** (2000) 383–408
15. Makinson, D., van der Torre, L.: Constraints for input-output logics. Journal of Philosophical Logic **30(2)** (2001) 155–185
16. Boella, G., van der Torre, L.: Rational norm creation: Attributing mental attitudes to normative systems, part 2. In: Procs. of ICAIL'03, New York (NJ), ACM Press (2003) 81–82
17. Hansson, B.: An analysis of some deontic logics. N*ôu*s **3** (1969) 373–398
18. Searle, J.: The Construction of Social Reality. The Free Press, New York (1995)
19. Hindriks, F.: The constitutive rule revisited. In: Procs. of 3rd Conference on Collective Intentionality, Rotterdam (2002)
20. Broersen, J., Dastani, M., Hulstijn, J., van der Torre, L.: Goal generation in the BOID architecture. Cognitive Science Quarterly **2(3-4)** (2002) 428–447
21. Lang, J., van der Torre, L., Weydert, E.: Utilitarian desires. Autonomous Agents and Multiagent Systems **5(3)** (2002) 329–363
22. Anderson, A.: The logic of norms. Logic et analyse **2** (1958)
23. Hart, H.: The Concept of Law. Clarendon Press, Oxford (1961)
24. Gelati, J., Governatori, G., Rotolo, N., Sartor, G.: Declarative power, representation, and mandate. A formal analysis. In: Procs. of JURIX 02, Amsterdam, IOS press (2002) 41–52
25. Castelfranchi, C., Tummolini, L.: The cognitive and behavioral mediation of institutions: Towards an account of institutional actions. In: Procs. of 4th Conference on Collective Intentionality. (2004)

26. Grossi, D., Dignum, F., Meyer, J.: Contextual terminologies. In: Procs. of CLIMA'05. (2005)
27. Grossi, D., Meyer, J., Dignum, F.: Modal logic investigations in the modal logic investigations in the semantics of counts semantics of counts-as as. In: Procs. of ICAIL'05, New York (NJ), ACM Press (2005) 1–9