A formal study on legal compliance and interpretation

Guido Boella
University of Turin

Guido Governatori
NICTA, Australia

Antonino Rotolo
University of Bologna

Leendert van der Torre
University of Luxembourg

Abstract

This paper proposes a logical framework to capture the norm change power and the limitations of the judicial system in revising the set of constitutive rules defining the concepts on which the applicability of norms is based. In particular, we reconstruct the legal arguments leading to an extensive or restrictive interpretation of norms.

1. Introduction and Motivation

An important research issue in AI is to design computer systems whose performance is constrained by suitable sets of legal norms: in this sense, norms establish what legality criteria should apply to their functioning (van der Torre, Boella, & Verhagen 2008). However, the general idea of regulating computer systems can be modelled in different ways. As, e.g., (Boella & van der Torre 2004) pointed out in the field of normative MAS, norms may work either as hard or soft constraints. In the first case, computer systems are designed in such a way as to avoid legal violations. In the second case, norms rather provide standards which can be violated, even though any violations should result in sanctions or other normative effects applying to non-compliant systems. In both perspectives, it is of paramount importance to develop mechanisms to enforcing and detecting *norm compliance*.

However, most logical models of legal reasoning often assume that norms give a complete description of their applicability conditions (see (Sartor 2005)). This view affects the concepts of compliance and violation. Indeed, if we assume that norms have a conditional structure such as $a_1, \ldots, a_n \Rightarrow Ob$ (if $a_1, \ldots, a_n$ hold, then b is obligatory), we are compliant with respect to this norm if b is obtained whenever we derive $a_1, \ldots, a_n$. However, the assumption that norms give complete description of their applicability conditions is too strong, due to the complexities and dynamics of the world. Norms cannot take into account all the possible conditions where they should or should not be applied, first of all because the legislator cannot consider all the possible contexts which are exceptional and he cannot foresee unexpected changes of the world (Hart 1958). Normative systems regulating real societies have two mechanisms to cope with this problem. First they distinguish regulative rules from constitutive rules. While the former, which are changed only by the legislative system, specify the ideal behaviour, the latter ones provide an ontology of institutional concepts to which the conditions of regulative rules refer to. Second, the judicial system is empowered to change the constitutive norms, under some restrictions not to go beyond the purpose from which the regulative rules stem. This combination of rules and norm change allows the legislator to disentangle the specification of a behaviour from the specific contexts of applicability. In this paper, we outline a logical framework which is able to capture the norm change power and at the same time the limitations of the judicial system in revising the set of constitutive rules defining the concepts on which the applicability of such a rule is based. Indeed, the distinction between regulative and constitutive rules (ontology vs norms) suggests that legal interpretation does not amount to revising norms, but to revising constitutive rules (Sartor 2006).

While the distinction between constitutive and regulative rules has been already introduced in fields such as MAS, the interpretation process has been only addressed in the field of AI and law but only as far as it concerns case based reasoning for common law (Bench-Capon 2002). Also the relation between constitutive norms and contexts has been considered (Grossi, Meyer, & Dignum 2008).

What is still lacking is a logical model of the interpretation mechanism which leads to dynamically revising constitutive norms to make the normative system flexible in handling concepts such those of violation and compliance. This issue breaks down into the following subquestions: How to model the meaning of an institutional concept? How to decide which constitutive norms to introduce to either shrink or extend the extension of the institutional concept? How to reason about the interaction between norms and goals?

Our methodology adopts an extension of Defeasible Logic (DL) (Governatori & Rotolo 2008a), which allows us to address these research issues.

2. Legal Rules and Legal Concepts

Checking legal compliance requires to establish if a legal rule r is violated by a fact or action f that happened under some circumstances c. Let us assume that r states that $\neg f$ ought to be the case. However, f is not necessarily a violation, because we also have to check whether c matches with the applicability conditions *App* of r (i.e., c implies *App*). In easy cases, this match and f directly amount to a violation. However, jurists argue that we have cases where this does not hold, as for example when there is a discrepancy between the literal meaning of r and its goal assigned by the legislator. If so, even though c matches with *App*, we do not have a violation because c should not match with *App*. A non-literal, goal-based interpretation of *App* would exclude c as a circumstance falling within the scope of r: *lex magis dixit quam voluit*, the law says more than what the legislator was meaning to say. Analogously, not all cases in which c mismatches with *App* are not violations. We could have that *lex minus dixit quam voluit*, the law says less than what the legislator was meaning to say: here a non-literal, goal-based interpretation of r would lead to broaden its applicability scope (Peczenik 1989).

To formally capture these scenarios we need to devise a reasoning framework consisting of the following components: a mechanism for reasoning about (i) legal concepts, (ii) legal rules, and (iii) the goal of legal rules.

The general idea behind this framework is that legal rules state what is obligatory and prohibited for the agents. In other words, they provide normative constraints for agent behaviour and we assume that no agent can change them (agents are not legislators). Legal concepts constitute the content of legal rules; in particular, they qualify their applicability conditions. Finally, as usually assumed in legal theory (Sartor 2005; Peczenik 1989), we assign goals to legal rules. In the social delegation cycle (Boella & van der Torre 2007) norms are planned starting from goals shared by the community of agents. However, such goals play also another role: they pose the limits within which the interpretation process of the judicial systems must stay when interpreting norms. We have two cases. First, a legal rule can be applied in a given situation, but if the norm were respected in that situation, the goal of the norm would be endangered by this. Second, a legal rule cannot be applied in a concrete case, but this situation leads to undermining the goal which such a rule is supposed to promote.

In both cases, an interpretation of the applicability of a norm by the judicial system is limited by the goal the legislator was aiming to when he devised the norm. Note that the goal alone would not be sufficient, since there could be many ways to achieve that goal. Thus, the norm works like a partial plan the legislator set up in advance. The judicial system is left with the task of dynamically adapt the applicability of the norm by revising the constitutive norms, in order to fulfil the goal of the norm also under unforeseen circumstances.

In this paper we adopt the view that legal concepts are built via constitutive rules having the so-called counts-as form (Searle 1995). For example, a common legal definition of holographic wills requires that they have been entirely handwritten and signed by the testator:

$$r_1 : \mathit{HandWritten\_Will}(x), \mathit{Signed\_Testator}(x) \Rightarrow_c \\ \Rightarrow_c \mathit{Holographic\_Will}(x)$$

This counts-as rule, if instantiated by any individual a, says that a counts as a holographic will if a has been entirely handwritten and signed by the testator.

Here, we will deal with such a type of constitutive rules following the approach described in (Governatori & Rotolo 2008b), where it is convincingly argued that an effective way to capture the basic properties of the counts-as link is to reframe it in terms of standard DL.

The set of legal rules (i.e., regulative rules) is kept to be fixed: any interpreter can argue about their applicability conditions but cannot either add new rules nor cancel them.

Legal rules will have the following form:

$$r_2 : \mathit{Vehicle}(x), \mathit{Park}(y) \Rightarrow_O \neg \mathit{Enter}(x,y)$$

This rule reads as follows: if x is a vehicle and y is a park, then it is forbidden for any x to enter y.

For the sake of simplicity, we will assume that legal rules only impose duties and prohibitions, and state permissions: they are captured within a suitable extension of standard DL (Governatori & Rotolo 2008a).

Finally, we define a set Goal of goals and a function $\mathscr{G}$ which maps legal rules into elements of Goal. For example, if $\mathscr{G}(r_2) = \textbf{road\_safety}$, this means that the goal of the rule prohibiting to enter into parks is to promote road safety[1]. The idea is quite standard in legal theory (Sartor 2006; Peczenik 1989; Sartor 2005) and has been already investigated in AI&Law, even though most works were mainly devoted to case-based reasoning and modeling case-law (Bench-Capon 2002). A similar idea has been recently proposed in the field of normative MAS by (Boella & van der Torre 2007), where it has been argued that norms derive from goals. In general, note that, for simplicity, goals are considered here as directly specified by the legal rules themselves, even though it is sometimes a hard task to determine what goals are supposed to be promoted by rules, a task which is usually accomplished by judges by developing suitable arguments during the trial.

3. Interpreting Legal Rules

Suppose Mary enters a park with her bike, thus apparently violating rule r_2 above about vehicles' circulation. Police stops her when she is still on her bike in the park and fines her. Mary thinks this is unreasonable and sues the municipality because she thinks that here the category "vehicle" should not cover bikes.

To establish if Mary violated r_2, we have two alternatives. The first is that the conceptual domain of the normative system, corresponding to a set of constitutive rules, allows for deriving that any bike b is indeed a vehicle:

$$\begin{aligned} T = \{ & r_3 : \mathit{2\_wheels}(x), \mathit{Transport}(x) \Rightarrow_c \mathit{Bike}(x), \\ & r_4 : \mathit{Bike}(x) \Rightarrow_c \mathit{Vehicle}(x) \} \end{aligned}$$

Alternatively, the conceptual domain could exclude that bikes are vehicles:

$$\begin{aligned} T' = \{ & r_3 : \mathit{2\_wheels}(x), \mathit{Transport}(x) \Rightarrow_c \mathit{Bike}(x), \\ & r_5 : \mathit{Bike}(x) \Rightarrow_c \neg \mathit{Vehicle}(x), \\ & r_6 : \mathit{Transport}(x) \rightsquigarrow_c \mathit{Vehicle}(x) \} \\ \succ = \{ & r_5 \succ r_6 \} \end{aligned}$$

[1] Hereafter, we will use bold type expressions to denote goals.

As usual in DL (Antoniou *et al.* 2001), our language includes (1) a superiority relation $\succ$ that establishes the relative strength of rules and is used to solve conflicts, (2) special rules marked with $\leadsto$, called defeaters, which are not meant to derive conclusions, but to provide reasons against the opposite. Indeed, T' also includes r_6, which states that, if we know that some x has purpose of transport, then we have reasons to block other rules which would lead to exclude that x is a vehicle. However, in T' r_6 is made weaker than r_5 via the superiority relation $\succ$, and so, if x is a bike, we conclude that x is not a vehicle.

Now, suppose the judge has to settle Mary's case. Here, the goal of legal rules such as r_2 may be decisive.

Indeed, if T is the case, the judge could argue that Mary should be fined, as r_2 clearly applies to her. But suppose that we can show that, if Mary's case fulfils the applicability conditions of r_2 (Mary's bike is a vehicle) then we should promote a goal which is incompatible with the goal assigned to r_2. For instance, if $\mathscr{G}(r_2) = \neg$**pollution**, prohibiting to circulate with bikes in parks would encourage people to get around parks by car and then walk. Hence, if we assume r_2 is fulfilled, this would be against the goal of r_2 and so the judge has good reasons to exclude that bikes are vehicles when r_2 should be applied. Accordingly, when arguing in this way, the judge may interpret r_2 by reducing its applicability conditions as far as Mary's case is concerned, and so by contracting T in order to obtain in T that Mary's bike is not a vehicle.

Suppose now that T' is the case. Here, the judge could argue that Mary should not be fined, as r_2 clearly does not apply. But suppose that, if r_2 is not fulfilled, this would be against the goal of r_2, which is now **pedestrian_safety**. In this case, the judge has rather good reasons to consider bikes as vehicles when r_2 is concerned. Hence, the judge may interpret r_2 by broadening its applicability conditions as far as Mary's case is concerned, and so by revising T' in such a way as Mary's bike is a vehicle.

In general, we should note that such types of revisions have to satisfy some requirements (let's still bear in mind the case of Mary's bike):

1. there is no other $g' \in$ Goal such that
   - the revision of T (or of T') promotes r_2's goal g which is incompatible, in the application context of r_2, with respect to the goal g' of another applicable rule r_3, and
   - $\mathscr{G}(r_2) \not\succ \mathscr{G}(r_3)$ ($\mathscr{G}(r_2)$ is not more important than $\mathscr{G}(r_3)$);
2. our set of constitutive rules should suggest us that the concept of *Bike* can be subsumed under the concept of *Vehicle*.

Point 1 above states that, if by contracting or revising the concept of *Bike*, we undermine at least one equally or more important goal, which is supposed to be promoted by another applicable rule, then such a contraction or revision is not acceptable. This limit is well-known by lawyers and legal theorists (Sartor 2005; Peczenik 1989), who often argue that any legal interpretation should be coherent within the legal system as a whole.

Point 2 above is rather connected with the fact that the set of constitutive rules should inherently provide some conceptual limits for any interpretation. Indeed, suppose that Mary enters the park with a gun. We could have reasons for arguing that entering with a gun is dangerous for all people in the park, and so for pedestrians too. However, this is not enough, of course, for arguing that guns are vehicle. In other words, if we do not have any other legal rules prohibiting to enter parks with guns, this behaviour will be permitted. Hence, point 2 has to do with Hart's (Hart 1958) theory of penumbra: we have a core of cases which can be clearly classified as belonging to the legal concept and a penumbra of hard cases, whole membership in the concept can be disputed; but hard cases should exhibit some conceptual link with the core of cases. This idea is formally captured here by confining the revision of the set of constitutive rules only to those situations where such a set, though failing to prove that a bike is a vehicle, already contains reasoning chains suggesting that this may be the case. For example, if we have

$$r_3 : \mathit{2\_wheels}(x), \mathit{Transport}(x) \Rightarrow_c \mathit{Bike}(x)$$
$$r_7 : \mathit{Bike}(x) \leadsto_c \mathit{Vehicle}(x)$$

r_7 states that, if we know that some x is a bike, this is not sufficient to prove that x is a vehicle (r_7 is a defeater), but it is sufficient to block other rules which would lead to exclude that x is a vehicle. This means that, possibly, if x is a bike, then it could not be unreasonable to consider x as a vehicle (for a similar reading of defeaters in terms of $\Diamond$, but applied to the concept of permission, see (Governatori & Rotolo 2008a)). Hence, the revision would require, for example, that r_7 is replaced by

$$r'_7 : \mathit{Bike}(x) \Rightarrow_c \mathit{Vehicle}(x)$$

The framework we have informally depicted above suggests that we also need a logical component to reason about rule goals. Such a component should enable us to check whether some situations promote rule goals or their negations. For our purpose it is sufficient to introduce a suitable set of rules for goals (Governatori & Rotolo 2008b) which should be used to establish what are the effects of situations where legal rules are violated or complied with, and, in doing so, to see whether they are consistent with the rule goals. In other words, we have to devise a set of rules like $d_1, \ldots, d_n \Rightarrow_G e$: if applicable in a given context D, this rule allows for deriving $G\,e$, meaning that e is a goal promoted by D. Consider once again rule r_2; suppose that its goal is **pedestrian_safety** and that Mary's case is described by the following set H of facts:

$$\begin{aligned} H = \{&\mathit{Bike}(b),\ \mathit{Park}(p),\ \mathit{Enter}(b,p) \\ &\mathit{NarrowSpace}(p),\ \mathit{UnprotectedChildrenArea}(p)\} \end{aligned}$$

H states that Mary enters the park p with her bike b. The park has narrow spaces for walking and there are unprotected children's play areas. This set assumes that r_2 is violated, at least in the hypothetical perspective in which Mary could not enter.

Suppose now that rules for goals correspond to the following set:

$$
\begin{aligned}
R^G = \{ & r_8 : \mathit{Bike}(x), \mathit{Park}(y), \mathit{Enter}(x,y) \Rightarrow_G \textbf{fast\_circulation} \\
& r_9 : \mathit{NarrowSpace}(x), \mathit{UnprotectedChildrenArea}(x), \\
& G\,\textbf{fast\_circulation} \Rightarrow_G \neg\textbf{pedestrian\_safety}\}
\end{aligned}
$$

Rule r_8 states that entering parks with bikes promotes the fast circulation of people in those parks; rule r_9 says that, if fast circulation is promoted and parks have narrow spaces and unprotected children's play areas, then the promoted goal is the negation of pedestrians safety. If so, if Mary's bike is allowed to enter, then we would promote a goal which is incompatible with the goal of r_2.

4. The Logical Framework

The following framework is an extension of DL; such an extension is line with works such as (Governatori & Rotolo 2008a; 2008b). In particular, on account of the informal presentation given in the previous section, while counts-as rules do not prove modalised literals, the system develops a constructive account of those modalities that rather correspond to obligations and goals: rules for these concepts are thus meant to devise suitable logical conditions for introducing modalities For example, while a counts-as rule such as $a_1, \ldots, a_n \Rightarrow_c b$, if applicable, will basically support the conclusion of b, rules such as $a_1, \ldots, a_n \Rightarrow_O b$ and $d_1, \ldots, d_n \Rightarrow_G e$ if applicable, will allow for deriving $O\,b$ and $G\,e$, meaning the former that b is obligatory, the latter that e is a goal promoted by the facts used to derive it (as previously explained).

Note that the framework is restricted to essentially propositional DL. Indeed, rules with free variables are interpreted as rule schemas, that is, as the set of all ground instances; in such cases we assume that the Herbrand universe is finite. This assumption is harmless in this context, as the rule applicability domains at hand always refer to finite set of individuals.

In our language, for $X \in \{c, O, G\}$, we have that $\phi_1, \ldots, \phi_n \rightarrow_X \psi$ is a *strict rule* such that whenever the premises $\phi_1, \ldots, \phi_n$ are indisputable so is the conclusion ψ. $\phi_1, \ldots, \phi_n \Rightarrow_X \psi$ is a *defeasible rule* that can be defeated by contrary evidence. A rule $\phi_1, \ldots, \phi_n \rightsquigarrow_X \psi$ is a *defeater* that is used to defeat some defeasible rules by supporting evidence to the contrary.

Definition 1 (Language) *Let* PROP *be a set of propositional atoms,* Goal *be a set of goal atoms,* $\mathrm{MOD} = \{c, O, G\}$, *and* Lbl *be a set of labels. The sets below are the smallest sets closed under the following rules:*

Literals and goals

$$
\begin{aligned}
\mathrm{Lit} &= \mathrm{PROP} \cup \{\neg p | p \in \mathrm{PROP}\} \\
\mathrm{GoalLit} &= \mathrm{Goal} \cup \{\neg g | g \in \mathrm{Goal}\}
\end{aligned}
$$

If q is a literal or a goal, $\sim q$ denotes the complementary literal or goal (if q is a positive literal or goal p then $\sim q$ is $\neg p$; and if q is $\neg p$, then $\sim q$ is p);

Modal literals and modal goals

$$
\begin{aligned}
\mathrm{ModLit} &= \{Xl, \neg Xl | l \in \mathrm{Lit}, X = O\} \\
\mathrm{ModGoal} &= \{Gg, \neg Gg | g \in \mathrm{GoalLit}\};
\end{aligned}
$$

Rules $\mathrm{Rul} = \mathrm{Rul_s} \cup \mathrm{Rul_d} \cup \mathrm{Rul_{dft}}$, *where* $X \in \{c, O\}$ *and* $\mathrm{Rul_s} = \mathrm{Rul_s^X} \cup \mathrm{Rul_s^G}$, $\mathrm{Rul_d} = \mathrm{Rul_d^X} \cup \mathrm{Rul_d^G}$, *and* $\mathrm{Rul_{dft}} = \mathrm{Rul_{dft}^X} \cup \mathrm{Rul_{dft}^G}$ *such that*

$$
\begin{aligned}
&\mathrm{Rul_s^X} = \{r : \phi_1, \ldots, \phi_n \rightarrow_X \psi | r \in \mathrm{Lbl}, A(r) \subseteq \mathrm{Lit}, \psi \in \mathrm{Lit}\} \\
&\mathrm{Rul_s^G} = \{r : \phi_1, \ldots, \phi_n \rightarrow_G \psi | \\
&\quad r \in \mathrm{Lbl}, A(r) \subseteq \mathrm{Lit} \cup \mathrm{ModLit} \cup \mathrm{ModGoal}, \psi \in \mathrm{GoalLit}\} \\
&\mathrm{Rul_d^X} = \{r : \phi_1, \ldots, \phi_n \Rightarrow_X \psi | r \in \mathrm{Lbl}, A(r) \subseteq \mathrm{Lit}, \psi \in \mathrm{Lit}\} \\
&\mathrm{Rul_d^G} = \{r : \phi_1, \ldots, \phi_n \Rightarrow_G \psi | \\
&\quad r \in \mathrm{Lbl}, A(r) \subseteq \mathrm{Lit} \cup \mathrm{ModLit} \cup \mathrm{ModGoal}, \psi \in \mathrm{GoalLit}\} \\
&\mathrm{Rul_{dft}^X} = \{r : \phi \rightsquigarrow_X \psi | r \in \mathrm{Lbl}, A(r) \subseteq \mathrm{Lit}, \psi \in \mathrm{Lit}\} \\
&\mathrm{Rul_{dft}^G} = \{r : \phi_1, \ldots, \phi_n \rightsquigarrow_G \psi | \\
&\quad r \in \mathrm{Lbl}, A(r) \subseteq \mathrm{Lit} \cup \mathrm{ModLit} \cup \mathrm{ModGoal}, \psi \in \mathrm{GoalLit}\}
\end{aligned}
$$

We use some obvious abbreviations, such as superscript for the rule mode (c, G, O), subscript for type of rule, and $\mathrm{Rul}[\phi]$ *for rules whose consequent is ϕ, for example:*

$$
\begin{aligned}
&\mathrm{Rul}^c = \{r : \phi_1, \ldots, \phi_n \hookrightarrow_c \psi | \hookrightarrow \in \{\rightarrow, \Rightarrow, \rightsquigarrow\}\} \\
&\mathrm{Rul_{sd}} = \{r : \phi_1, \ldots, \phi_n \hookrightarrow_X \psi | X \in \mathrm{MOD}, \hookrightarrow \in \{\rightarrow, \Rightarrow\}\} \\
&\mathrm{Rul_s}[\psi] = \{\phi_1, \ldots, \phi_n \rightarrow_X \psi | X \in \mathrm{MOD}\}
\end{aligned}
$$

We use $A(r)$ to denote the set $\{\phi_1, \ldots, \phi_n\}$ of antecedents *of the rule r, and $C(r)$ to denote the* consequent *ψ of the rule r.*

Let us now introduce the notion of normative theory, which is the knowledge base used to reason about the applicability of legal rules.

Definition 2 (Normative Theory) *A normative theory is a structure*

$$D = (F, G, R^c, R^O, R^G, \succ, \mathscr{G}, >)$$

where

- $F \subseteq \mathrm{Lit} \cup \mathrm{ModLit} \cup \mathrm{ModGoal}$ *is a finite set of facts;*
- $G \subseteq \mathrm{GoalLit}$ *is a set of rule goals,*
- $R^c \subseteq \mathrm{Rul}^c$ *is a finite set of counts-as rules,*
- $R^O \subseteq \mathrm{Rul}^O$ *is a finite set of obligation rules,*
- $R^G \subseteq \mathrm{Rul}^G$ *is a finite set of goal rules,*
- $\succ$ *is an acyclic superiority relation defined over* $R^c \times R^c \cup R^O \times R^O \cup R^G \times R^G$,
- $\mathscr{G} : R^O \mapsto G$ *is a function assigning a goal to each obligation rule,*
- $>$ *is a partial order over G defining the relative importance of the rule goals.*

Proofs are sequences of literals and modal literals together with so-called proof tags $+\Delta$, $-\Delta$, $+\partial$ and $-\partial$. If $X \in \{c, O, G\}$, given a normative theory D, $+\Delta^X q$ means that literal q is provable in D using only facts and strict rules for X, $-\Delta^X q$ means that it has been proved in D that q is not definitely provable in D, $+\partial^X q$ means that q is defeasibly provable in D, and $-\partial^X q$ means that it has been proved in D that q is not defeasibly provable in D.

Definition 3 *Given a normative theory D, a proof in D is a linear derivation, i.e, a sequence of labelled formulas of the type $+\Delta^X q$, $-\Delta^X q$, $+\partial^X q$ and $-\partial^X q$, where the proof conditions defined in the rest of this section hold.*

Definition 4 *Let D be a normative theory. Let $\# \in \{\Delta, \partial\}$ and $X \in \{O, G\}$, and $P = (P(1), \dots, P(n))$ be a proof in D. A literal q is* #-provable *in P if there is a line $P(m)$, $1 \leq m \leq n$, of P such that either*

1. *q is a literal and $P(m) = +\#^c q$ or*
2. *q is a modal literal or a modal goal Xp and $P(m) = +\#^X p$ or*
3. *q is a modal literal or a modal goal $\neg Xp$ and $P(m) = -\#^X p$.*

A literal q is #-rejected *in P if there is a line $P(m)$ of P such that*

1. *q is a literal and $P(m) = -\#^c q$ or*
2. *q is a modal literal or a modal goal Xp and $P(m) = -\#^X p$ or*
3. *q is a modal literal or a modal goal $\neg Xp$ and $P(m) = +\#^X p$.*

The definition of Δ^X, $X \in \{c, O, G\}$ describes just forward (monotonic) chaining of strict rules[2]:

$+\Delta^X$: If $P(n+1) = +\Delta^X q$ then
(1) $q \in F$ if $X = c$ or $Xq \in F$ or
(2) $\exists r \in R_s^X[q] : \ \forall a \in A(r)\ a$ is Δ-provable.

Instead, to reason defeasibly and solve rule conflicts, we need to state when one rule is stronger than another:

Definition 5 *Let $D = (F, G, R^c, R^O, R^G, \succ, \mathscr{G}, >)$ be a normative theory. A rule r* prevails *over another rules s iff*

- *$\mathscr{G}(r) > \mathscr{G}(s)$ or*
- *$r \prec s$ and $\mathscr{G}(s) \not> \mathscr{G}(r)$*

To show that a literal q is defeasibly provable (see the proof conditions below) with the mode X we have two choices: (1) We show that q is already definitely provable; or (2) we need to argue using the defeasible part of a normative theory D. For this second case, some (sub)conditions must be satisfied. First, we need to consider possible reasoning chains in support of $\sim q$ with the mode X, and show that $\sim q$ is not definitely provable with that mode (2.1 below). Second, we require that there must be a strict or defeasible rule with mode X for q which can be applied (2.2 below). Third, we must consider the set of all rules which are not known to be inapplicable and which permit to get $\sim q$ with the mode X (2.3 below). Essentially, each such a rule s attacks the conclusion q. For q to be provable, s must be counterattacked by a rule t for q with the following properties: (i) t must be applicable, and (ii) t must prevail over s. Thus each attack on the conclusion q must be counterattacked by a stronger rule. In other words, r and the rules t form a team (for q) that defeats the rules s. Note that in our framework, in addition to $\succ$, also goals can be used to determine the relative strength of any legal rule in case of conflicts with other rules; the following definition enables us to handle together goal preferences and the superiority relation $\succ$.

$+\partial^X$: If $P(n+1) = +\partial^X q$ then
(1)$+\Delta^X q \in P(1..n)$ or
(2) (2.1) $-\Delta^X \sim q \in P(1..n)$ and
  (2.2) $\exists r \in R_{\mathrm{sd}}^X[q]$ such that $\forall a \in A(r)\ a$ is ∂-provable, and
  (2.3) $\forall s \in R^X[\sim q]$ either $\exists a \in A(s)$ such that a is ∂-rejected, or
    (2.3.1) $\exists t \in R^X[q]$ such that $\forall a \in A(r)\ a$ is ∂-provable and
      t prevails over s

Definition 6 *Given a normative theory D, $D \vdash \pm\#^X l$ (i.e., $\pm\#^X l$ is a conclusion of D), where $\# \in \{\Delta, \partial\}$ and $X \in \{c, O, G\}$, iff there is a proof $P = (P(1), \dots, P(n))$ in D such that $P(n) = \pm\#^X l$.*

It is worth noting that our logic enjoys nice computational properties:

Theorem 1 *For every normative theory D, the conclusions of D can be computed in time linear to the size of the theory, i.e., $O(|U^D| * |R|)$, where U^D the set of all the atoms and atomic goals occurring in D.*

Proof The proof comes directly from the result provided in (Governatori & Rotolo 2008b; 2008a). In fact, the current logic is structurally similar to those presented there.

5. Interpretation: Revising Constitutive Rules

Let us consider a normative theory D and a legal rule $r_0 : b_1, \dots, b_n \hookrightarrow_O l$ in it such that the goal of r_0 is g. As informally discussed before, when we have to assess if $\sim l$ and a certain set of circumstances H amount to a violation of r_0 we have to consider two cases. First, we have $\neg l$, which, under the circumstances H, undermines the goal g of r_0; however, if the agent did l, this action would still produce $\neg g$, thus supporting the view that the case should be excluded from the applicability range of r_0 even though H matches with $b_1, \dots, b_n$ (*Lex magis dixit quam voluit*). Second, we have $\neg l$, which, under the circumstances H, undermines the goal g of r_0; however, if the agent did l, this action would produce g, thus supporting the view that the case should be included in the applicability range of r_0 even though H does not match with $b_1, \dots, b_n$ (*Lex minus dixit quam voluit*).

The discussion above requires to formally characterise those situations where a context and an action or fact fulfil a norm and those situations where we undermine the goal of a norm.

Definition 7 (Rule Fulfilment) *Let a context be a set $H = \{f_1, \dots, f_m\}$ of literals. A normative theory*

$$D = (F, G, R^c, R^O, R^G, \succ, \mathscr{G}, >)$$

and H fulfil $r_0 \in R_{\mathrm{sd}}^O$ iff, $\forall b_t \in A(r_0)$,

$$D' = (F \cup H, G, R^c, R^O, R^G, \succ, \mathscr{G}, >) \vdash +\partial^O C(r_0)$$
$$D' = (F \cup H, G, R^c, R^O, R^G, \succ, \mathscr{G}, >) \vdash +\partial^c b_t$$

and

- *$D' \vdash -\partial^c l$ or $D' \vdash +\partial^c \neg l$ when $C(r_0)$ is a negative literal $\neg l$ (r_0 is a conditional prohibition);*
- *$D' \vdash +\partial^c l$ when $C(r_0)$ is a positive literal l (r_0 is a conditional obligation).*

[2] For space reasons, we omit the proof conditions for $-\Delta$ and $-\partial$. See (Governatori & Rotolo 2008a) for the method to obtain them.

Definition 8 (Goal Violation) *A normative theory*

$$D = (F, G, R^c, R^O, R^G, \succ, \mathscr{G}, >)$$

and the context $H = \{f_1, \ldots, f_m\} \cup \{\sim l\}$ *violate the goal* g *of* $r_0 : a_1, \ldots, a_n \hookrightarrow_O l \in R^O_{\mathrm{sd}}$ *iff*

$$\begin{aligned}&(F, G, R^c, R^O, R^G, \succ, \mathscr{G}, >) \vdash -\partial^G \neg g_0\\&(F \cup H \cup \{\sim l\}, G, R^c, R^O, R^G, \succ, \mathscr{G}, >) \vdash +\partial^G \neg g_0\end{aligned}$$

Definition 9 *Let* $D = (F, G, R^c, R^O, R^G, \succ, \mathscr{G}, >)$ *be a normative theory. A counts-as reasoning chain* $\mathscr{C}$ *in* D *for a literal* l *is a finite sequence* $\mathscr{R}_1, \ldots, \mathscr{R}_n$ *where*

- $\mathscr{R}_i \subseteq R^c$, $1 \leq i \leq n$,
- $\mathscr{R}_n = \{a_1, \ldots, a_n \hookrightarrow_c l | \hookrightarrow \in \{\rightarrow, \Rightarrow, \leadsto\}\}$,
- $\mathscr{R}_k \subseteq R^c$, $1 < k \leq n$, *is such that* $\forall\, r^k \in \mathscr{R}_k$, $\forall b \in A(r^k)$: $\exists r^{k-1} \in \mathscr{R}_{k-1} : b = C(r^{k-1})$.

For all $s \in \mathscr{R}_i$, $1 \leq i \leq n$, *we will say that* s *is in* $\mathscr{C}$. *If a literal* p *occurs in the head or the body of any* s *in* $\mathscr{C}$, *we will say that* p *is in* $\mathscr{C}$. *We define analogously a goal or an obligation reasoning chain* $\mathscr{C}$ *in* D *for a literal* l *when all rules in* $\mathscr{C}$ *are in* R^G *or* R^O, *respectively.*

We are now ready to formally define the operations of expansion (*Lex minus dixit quam voluit*) and contraction (*Lex magis dixit quam voluit*) of the applicability conditions of a norm.

Definition 10 (Rule Expansion) *Let*

$$D = (F, G, R^c, R^O, R^G, \succ, \mathscr{G}, >)$$

be a normative theory, $r_0 : b_1, \ldots, b_n \hookrightarrow_O l \in R^O$ *be a regulative rule, and* $H = \{f_1, \ldots, f_m\}$ *be a context. If*

1. $\{b_k, \ldots b_{k+j}\} \subseteq A(r_0)$ *and, for all* b_t, $k \leq t \leq k+j$,
$$(F \cup H, G, R^c, R^O, R^G, \succ, \mathscr{G}, >) \not\vdash +\partial^c b_t$$
2. *D and H violate the goal g of* r_0, *and*
3. *there exist the counts-as reasoning chains* $\mathscr{C}_k, \ldots \mathscr{C}_{k+j}$ *in* D *for* $b_k, \ldots b_{k+j}$, *such that for each* $f \in \{f_1, \ldots, f_m\}$, f *is in* $\mathscr{C}_h$, $k \leq h \leq k+j$,

then the expansion of the applicability conditions of a regulative rule r_0 *with respect to the context* H *corresponds to the following operation* $D^*_{b_{k,\ldots,b_{k+j}}}$ *over* D:

$$D^*_{b_{k,\ldots,b_{k+j}}} = (F, G, R'^c, R^O, R^G, \succ', \mathscr{G}, >)$$

where

$$\begin{aligned}R'^c =& R^c - \{r' : d_1, \ldots, d_n \leadsto_c e | r' \textit{ is in } \mathscr{C}_h\} \cup\\ &\cup \{r' : d_1, \ldots, d_n \Rightarrow_c e\}\\ \succ' =& (\succ \cup \{r' \succ s | r' \textit{ is in } \mathscr{C}_h, s \in R^c[\sim C(r')]\}) -\\ &- \{t \succ r' | t \in R^c[\sim C(r')]\}\end{aligned}$$

- *such that* $D' = (F \cup H, G, R'^c, R^O, R^G, \succ', \mathscr{G}, >) \vdash -\partial_G \neg g'$, *where* g' *is the goal of any rule* $z \in R^O_{\mathrm{sd}}$ *applicable in* D' *such that* $g \not> g'$.

Definition 11 (Rule Contraction) *Let*

$$D = (F, G, R^c, R^O, R^G, \succ, \mathscr{G}, >)$$

be a normative theory, $r_0 : b_1, \ldots, b_n \hookrightarrow_O l \in R^O$ *be a regulative rule, and* $H = \{f_1, \ldots, f_m\}$ *be a context. If*

1. *D and H fulfil* r_0,
2. *D and H violate the goal g of* r_0,
3. $b_k \in A(r_0)$, *such that* b_k *occurs in every goal reasoning chain* $\mathscr{C}$ *for* $\neg g$ *in the normative theory*
$$(F \cup H, G, R^c, R^O, R^G, \succ, \mathscr{G}, >)$$

then the contraction of the applicability conditions of a regulative rule r_0 *with respect to the context* H *corresponds to the following operation* $D^-_{b_k}$ *over* D:

$$D^-_{b_k} = (F, G, R'^c, R^O, R^G, \succ', \mathscr{G}, >)$$

where

$$\begin{aligned}R'^c =& R^c \cup \{r : f_1, \ldots, f_m \leadsto\ \sim b_k\}\\ \succ' =&\ \succ - \{s \succ r \mid r \in R'^c - R^c\}.\end{aligned}$$

- *such that* $D' = (F \cup H, G, R'^c, R^O, R^G, \succ', \mathscr{G}, >) \vdash -\partial_G \neg g'$, *where* g' *is the goal of any rule* $z \in R^O_{\mathrm{sd}}$ *applicable in* D' *such that* $g \not> g'$.

Example 1 *Consider the following normative theory augmented with the context H regarding Mary's case:*

$$\begin{aligned}F =& \{\textit{Park},\ \textit{UnprotectedChildrenArea},\ \textit{NarrowSpace}\}\\ H =& \{\textit{2\_wheels}, \textit{Transport},\ \neg\textit{Engine}\}\\ G =& \{\textbf{fast\_circulation},\ \textbf{pedestrian\_safety}\}\\ R^c =& \{r_3 : \textit{2\_wheels}, \textit{Transport} \Rightarrow_c \textit{Bike}\\ & r_4 : \textit{Bike} \leadsto_c \textit{Vehicle}\\ & r_{10} : \textit{Transport}, \neg\textit{Engine} \Rightarrow_c \neg\textit{Vehicle}\}\\ R^O =& \{r_2 : \textit{Vehicle}, \textit{Park} \Rightarrow_O \neg\textit{Enter},\\ & r_{11} : \textit{Vehicle}, \textit{NarrowSpace} \Rightarrow_O \neg\textit{Stop}\}\\ R^G =& \{r_8 : \textit{Bike}, \textit{Park}, \textit{Enter} \Rightarrow_G \textbf{fast\_circulation}\\ & r_9 : \textit{NarrowSpace}(x), \textit{UnprotectedChildrenArea}(x),\\ & \qquad G\textbf{fast\_circulation} \Rightarrow_G \neg\textbf{pedestrian\_safety}\\ & r_{12} : \textit{NarrowSpace}, \textit{Vehicle} \Rightarrow_G \textbf{fast\_circulation}\\ & r_{13} : \textit{Bike}, \textit{Park}, \neg\textit{Enter} \Rightarrow_G \textbf{pedestrian\_safety}\}\\ \succ =& \{r_{10} \succ r_4\}\\ \mathscr{G} =& \{\mathscr{G}(r_2) = \textbf{pedestrian\_safety},\\ & \mathscr{G}(r_{11}) = \textbf{fast\_circulation}\}\\ > =& \{\textbf{pedestrian\_safety} > \textbf{fast\_circulation}\}\end{aligned}$$

Suppose Enter holds. This may correspond to a potential violation of r_2. *This is not the case, because* r_2 *is not triggered and we do not derive Vehicle. However, we obtain* $\neg g$ *via* r_9, *i.e., we undermine the goal of* r_2. *Since we have* r_4 *we can construct a counts-as reasoning chain supporting Bike, and so we can expand the applicability conditions of* r_2. *Via* r_{13} *we also promote the goal of* r_2. *Doing so, we trigger* r_{12} *and promote* **fast_circulation**, *which is incompatible with* **pedestrian**$_{\mathbf{s}}$**afety** *(via* r_9*). However,* **pedestrian_safety** *is more important than* **fast_circulation**.

Now, suppose that r_4 is

$$Bike \Rightarrow_c Vehicle$$

and **pedestrian_safety** *is less important than* **fast_circulation**. *We can contract the applicability conditions of r_2 by adding a defeater*

$$2\_wheels, Transport \neg Engine \rightsquigarrow_c \neg Bike$$

and change the superiority to make this defeater stronger than r_3, thus satisfying Definition 11's requirements.

Note that the operations $D^*_{b_k,\dots,b_{k+j}}$ and $D^-_{b_k}$ introduced in Definitions 10 and 11 correspond to special cases of AGM revision and contraction of conclusions in DL (Billington *et al.* 1999). Indeed, under some preconditions, expanding the applicability conditions of a norm amounts to modifying the rules and the superiority relation even if the negation of one or more elements in $b_k, \dots, b_{k+j}$ are derivable in D. However, due to the sceptical nature of DL, we still do not get a contradiction. On the other hand, under suitable preconditions, contracting the applicability conditions of a norm corresponds to preventing the proof of b_k. R'^c ensures that if b_k has been proven, a defeater with head $\neg b_k$ will fire. (Billington *et al.* 1999) provided a reformulation within DL of AGM postulates for revision and contraction: the results provided there can be extended to our framework

Theorem 2 *If preconditions 1, 2, 3 of Definition 10 hold, $D^*_{b_k,\dots,b_{k+j}}$ satisfies (Billington* et al. *1999)'s reformulation of AGM postulates for revision. If preconditions 1, 2, 3 of Definition 11 hold, $D^-_{b_k}$ satisfies (Billington* et al. *1999)'s reformulation of AGM postulates for contraction.*

6. Related Work and Conclusions

An extensive literature is devoted to legal ontologies (see, e.g., the survey in (Casanovas 2008)), but it is oriented to develop applications in the field of semantic web and rule interchange languages for the legal domain, applications which are not our primary concern. Also, these approaches usually fail to deal with the defeasibility and dynamics of legal concepts. The possibility to model legal and normative ontologies via constitutive rules has a solid philosophical backgrounds (see (Searle 1995; Sartor 2006)). However, to the best of our knowledge, there is no work so far devoted to the dynamics and revision of constitutive rules, and no proposal regarding how to model the interpretation of regulative rules in these terms. The only work which addressed the problem of the penumbra of legal concepts within a complete theory of counts-as rules is (Grossi 2007). (Grossi 2007) provides very complex modal account of counts-as rules. But what is lacking in that work, too, is that it does not address the problem of the dynamics of constitutive rules and does not consider the role of normative goals in determining the applicability conditions of regulative rules. Thus, we believe that this paper may indeed contribute to fill a gap in the literature, as it is almost standard in legal theory the idea that the goals of regulative rules are decisive in clarifying the scope of the legal concepts that qualify the applicability conditions for those rules.

Acknowledgements

This paper extends (Boella *et al.* 2010b) and is an earlier version of (Boella *et al.* 2010a). NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy, the Australian Research Council through the ICT Centre of Excellence program and the Queensland Government.

References

Antoniou, G.; Billington, D.; Governatori, G.; and Maher, M. J. 2001. Representation results for defeasible logic. *ACM Transactions on Computational Logic* 2(2):255–287.

Bench-Capon, T. J. M. 2002. The missing link revisited: The role of teleology in representing legal argument. *Artif. Intell. Law* 10(1-3):79–94.

Billington, D.; Antoniou, G.; Governatori, G.; and Maher, M. 1999. Revising nonmonotonic belief sets: The case of defeasible logic. In *Proc. KI-99*. Springer.

Boella, G., and van der Torre, L. 2004. Fulfilling or violating obligations in multiagent systems. In *Procs. IAT04*.

Boella, G., and van der Torre, L. 2007. Norm negotiation in multiagent systems. *Int. Journal Coop. Inf. Syst.* 16(1).

Boella, G.; Governatori, G.; Rotolo, A.; and van der Torre, L. 2010a. *Lex minus dixit quam voluit, lex magis dixit quam voluit*: A formal study on legal compliance and interpretation. In *Proc. AICOL 2009*. Springer.

Boella, G.; Governatori, G.; Rotolo, A.; and van der Torre, L. 2010b. A logical understanding of legal interpretation. In *Proc. KR 2010*. AAAI.

Casanovas, P., ed. 2008. *Proc. LOAIT 2007*. CEUR.

Governatori, G., and Rotolo, A. 2008a. Bio logical agents: Norms, beliefs, intentions in defeasible logic. *Autonomous Agents and Multi-Agent Systems* 17(1):36–69.

Governatori, G., and Rotolo, A. 2008b. A computational framework for institutional agency. *Artif. Intell. Law* 16(1):25–52.

Grossi, D.; Meyer, J.; and Dignum, F. 2008. The many faces of counts-as: A formal analysis of constitutive rules. *J. Applied Logic* 6(2):192–217.

Grossi, D. 2007. *Desigining Invisible Hancuffs. Formal Investigations in Institutions and Organizations for Multi-Agent Systems*. Ph.D. Dissertation, Utrecht University.

Hart, H. 1958. Positivism and the separation of law and morals. *Harvard Law Review* 71(4):593–629.

Peczenik, A. 1989. *On law and reason*. Kluwer.

Sartor, G. 2005. *Legal reasoning: A cognitive approach to the law*. Springer.

Sartor, G. 2006. Fundamental legal concepts: A formal and teleological characterisation. *Artif. Intell. Law* 14(1-2):101–142.

Searle, J. 1995. *The Construction of Social Reality*. New York: The Free Press.

van der Torre, L.; Boella, G.; and Verhagen, H., eds. 2008. *Normative Multi-agent Systems*, Special Issue of *JAAMAS*, vol. 17(1).