Reasoning with Normative Systems

Giovanni Sartor

Abstract The cognitive attitudes and operations involved in dealing with large normative systems are significantly different from those involved in complying with isolated social norms. While isolated norms may be directly applied by the agents endorsing them, this does not happen with regard to large normative systems. In the latter case, the agent must first inquire what the system requires from him (or what it allows him to do), namely, what is obligatory or permitted with regard to the normative system, and thus what would be required for complying with it, under different circumstances. I shall propose an argumentation-based approach for enabling an agent to process such requests, as resulting from a normative system and the existing factual circumstances.

Keywords Normative systems • Deontic logic • Legal reasoning • Dynamics of legal systems

1 Introduction

Human and artificial agents take into account not only shared social norms, but also complex institutional systems. We are often faced with systems of this kind in our daily life (the legal system, but also the prescriptions of an institutionalised religion, or the regulations of a company, a condominium, a regulated market, a teaching institution, a sociotechnical infrastructure such as an airport or a harbour, etc.). Most norms in such systems are created by norm-creating acts of the regulators of such systems (public or private authorities), and their content is to a large extent

---

Some parts of this paper have been published in the volume "The Goals of Cognition. Essays in Honor of Cristiano Castelfranchi".

G. Sartor (✉)
Law Faculty – CIRSFID, University of Bologna, Bologna, Italy

European University Institute of Florence, Florence, Italy
e-mail: giovanni.sartor@gmail.com

A. Herzig, E. Lorini (eds.), *The Cognitive Foundations of Group Attitudes and Social Interaction*, Studies in the Philosophy of Sociality 5,
DOI 10.1007/978-3-319-21732-1_9

dependent upon the discretionary choice of the regulator. Moreover such norms regulate very specific and differentiated situations, with which most agents do not have previous acquaintance. Thus the precise content of such norms cannot be derived from shared values and attitudes, nor can be induced from social behaviour. Moreover, an institutional system can contain a huge number of such norms, up to hundreds of thousand, so that it exceeds the storage capacity of the human mind. Such norms are also subject to frequent change and to multiple interpretations, so that even if they could be stored in a single repository, the repository would soon become useless unless continuously updated.

This means that for bounded agents, the way of learning the content of a complex institutional system must differ from the way of learning social norms.

When we learn social norms we permanently store them in our memory, as the content of appropriate normative beliefs and goals, so that they can directly govern our behaviour. On the contrary, we do cannot learn and store in our memory most norms included in large normative systems. We rather possess some ideas about the existence of such a system and the ways to identify its content. When needed, we collect some fragmentary information about the system and combine this information with the relevant facts, both tasks being often delegated to experts. On the basis of this information we can conclude that the system requires us to perform certain actions.

When referring to a large normative system *N* an agent usually does not immediately find an answer to the question "What ought I to do?" (as it usually happens when applying a shared social norm the agent is endorsing). The agent rather needs to asks itself (or the appropriate expert) "What does *N* require from me?," i.e., "What ought I do to according to *N*?" Agents may have to deal with different normative systems and distinguish the requests provided by each one of them.

I will propose an argumentation approach, which takes a normative system and the relevant facts as inputs, in order to deliver such answers.

2 Preliminary Notions: Actions, Obligations, Norms

For reasoning with normative systems, we need some basic notions. First, a way of expressing action and obligations is required. For actions I will use the simple *E* operator of Pörn (1977) (on the *E* operator see also Sergot 2001), though other action logics, such as STIT (Belnap et al. 2001), would be appropriate as well.

Definition 1 (Actions). Let proposition $\mathsf{E}_j\phi$ describe agent *j*'s action consisting in the production of state of affairs ϕ, where "ϕ" is any proposition. Thus $\mathsf{E}_j\phi$ means "*j* brings it about that ϕ". The non-accomplishment of an action is therefore described by $\neg\mathsf{E}_j\phi$, i.e., *j* does not bring about that *A*.

For simplicity, when an agent brings about its own action, I will not repeat the agent's name in the action's result. Thus, for expressing the idea that *John* smokes

(*John* brings it about that he smokes) rather than writing $\mathsf{E}_{John}Smoke(John)$, I will write $\mathsf{E}_{John}Smoke$.

As an example of an action-proposition, consider the following

$$\mathsf{E}_{John}Damaged(Tom)$$

which means "*John* brings it about that *Tom* is damaged", or more simply "*John* damages *Tom*" while the following

$$\neg\mathsf{E}_{John}Damaged(Tom)$$

means "*John* does not bring it about that *Tom* is damaged", or "*John* does not damage *Tom*". I shall adopt the logic of E, which is a classical modal logic (if A and B are logically equivalent, then $\mathsf{E}_xA \leftrightarrow \mathsf{E}_xB$) including the axiom schema:

$$\mathsf{E}_x\phi \Rightarrow \phi$$

meaning that if the state of affairs ϕ is realised though an action, then it is the case that ϕ. For instance, the fact that *Tom* makes it so that *Ann* suffers damage, obviously entails that *Ann* suffers damage:

$$\mathsf{E}_{Tom}Damaged(Ann) \Rightarrow Damaged(Ann)$$

Besides an action logic E, I need a deontic logic to express obligations.

Definition 2 (Obligations and prohibitions). Let O denote obligation. $O\mathsf{E}_j\phi$ means "it is obligatory that j brings it about that ϕ". Similarly $O\neg\mathsf{E}_j\phi$ means "it is obligatory that j does not bring about that ϕ", or "it is forbidden that j brings about that ϕ".

For instance, the following means "it is obligatory that *John* makes it so that *Tom* is compensated", or more simply, "it is obligatory that *John* compensates *Tom*",

$$O\mathsf{E}_{John}Compensated(Tom)$$

while the following means "it is forbidden that *John* damages *Tom*".

$$O\neg\mathsf{E}_{John}Damages(Tom)$$

As usual, I take permission to be the negation of prohibitions. I will not endorse here a particular deontic logic, since the following considerations may apply to different deontic logics. The reader may assume, for instance, standard deontic logic, which is a normal modal logic including, besides all tautologies of propositional logic, definition **Df P**.$P\phi \leftrightarrow \neg O\neg\phi$, axiom **K**.$O(\phi \Rightarrow \psi) \Rightarrow (O\phi \Rightarrow O\psi)$, axiom **D**.$O\phi \Rightarrow P\phi$ and the necessitation rule, **N** according to which if ϕ is a theorem, so is $O\phi$. We may on the other hand distinguish obligations directly established by the norms in the normative system of the system and derived obligation extracted from such norms, e.g., indicating what is necessary for complying with them (van der Torre and Hansen 2008). This too would be consistent with our framework, but we cannot explore it here.

For representing legal contents, we need norms, which can be viewed a kind of defeasible conditional.

Definition 3 (Norm). A *norm* has the form

$$A \Rightarrow B$$

where A is the antecedent condition, B the ensuing normative conclusion, and $\Rightarrow$ expresses a defeasible unidirectional connection, according to which antecedent A triggers conclusion B. In the norm the antecedent A is a proposition and consequent B is any kind of deontic or constitutive normative qualification.

Thus, a norm $A \Rightarrow B$ captures the unidirectional defeasible connection between an antecedent (possibly empty) fact and the normative consequent that is generated by that fact: normative effect B is triggered when the antecedent condition A holds. We write $A \not\Rightarrow B$ for the statement that the norm's antecedent fails to support the norm's conclusion, so that the norm cannot be applied in valid inferences. Arguments establishing that $A \not\Rightarrow B$ undercut the use of the norm $A \Rightarrow B$ in valid inferences.

Here is an example of two deontic norms, the first stating that it is forbidden to cause damage to others, and the second that who causes damage to another has the obligation to compensate the latter (in the following when obvious I drop the requirement $x \neq y$):

$$x \neq y \Rightarrow O\neg \mathsf{E}_x Damaged(y)$$

$$x \neq y \wedge \mathsf{E}_x Damaged(y) \Rightarrow O\mathsf{E}_x Compensated(y)$$

The following is an example of a constitutive norm, saying that if we injure a person (make so that someone is injured), we cause damage to that person (injuring counts as damaging):

$$\mathsf{E}_x Injured(y) \Rightarrow \mathsf{E}_x Damaged(y)$$

Note that I do not distinguish deontic conditionals and constitutive or counts-as conditionals, since both are modelled as defeasible conditionals (Searle 1995; Jones and Sergot 1996).

I assume an argumentation system as defined in Prakken (2010). Such a system includes two sets of inference rules, strict and defeasible inference rules, which have to be applied to a knowledge base of premises.

Strict inference rules have the form $[\phi_1, \ldots \phi_n] \mapsto \psi$. The conclusion ψ of the strict rule holds without exceptions when all its antecedent conditions $\phi_1, \ldots \phi_n$ hold; therefore the application of the rule to derive ψ cannot be challenged unless at least one antecedent condition in $\phi_1, \ldots \phi_n$ is also challenged.

Defeasible inference rules have the form $[\phi_1, \ldots \phi_n] \rightsquigarrow \psi$; the conclusion ψ of the defeasible rule holds only presumptively (with the possibility of exceptions) when all its antecedent conditions $\phi_1, \ldots \phi_n$ hold; therefore the application of the rule to derive ψ can be challenged also without challenging the antecedent conditions, i.e. by rebutting the rule's conclusion or by undercutting the rule's application.

Rules of both kinds can applied to a knowledge base of premises, these being formulas in a logical language.

Here I shall just introduce the main idea of an argumentation system in an informal way. The model I propose is inspired by Prakken (2010), to which I refer for a detailed presentation, though my account will depart from it in some aspects, to provide a simpler framework.

Premises, i.e., formulas in the underlying logical language $\mathscr{L}$ are basic arguments. Further arguments can constructed by applying inference rules to the conclusions of arguments already available: thus given arguments $A_1, \ldots A_n$ with conclusion $\phi_1, \ldots \phi_n$, through an inference rule $[\phi_1, \ldots \phi_n] \mapsto \psi$ we can obtain argument $B_s = \{A_1, \ldots A_n \mapsto \psi\}$, while through an inference rule $[\phi_1, \ldots \phi_n] \rightsquigarrow \psi$ we can obtain argument $B_d = \{A_1, \ldots A_n \rightsquigarrow \chi\}$. For instance, given premises a and $a \Rightarrow b$ and inference rule $[\phi, \phi \Rightarrow \psi] \rightsquigarrow \psi$, we can construct arguments $A_1 = \{a\}$, $A_2 = \{a \Rightarrow b\}$, and $A_3 = \{A_1, A_2 \rightsquigarrow b\}$, i.e., $\{\{a\}, \{a \Rightarrow b\} \rightsquigarrow b\}$.

Arguments may be defeated (rebutted or undercut) by counterarguments: rebutting takes place when an argument having a conclusion ψ through a defeasible rule (as its ultimate conclusion, or the conclusion of one of its subarguments) faces a non weaker counterargument having the complementary conclusion $\overline{\psi}$; undercutting takes place when an argument including a defeasible rule $[\phi_1, \ldots \phi_n] \rightsquigarrow \psi$, having name r (we assume that each rule has an unique name) has a counterargument with conclusion $\neg r$ (the negation of a rule-name being understood as the denial of the rule's applicability). An argument is justified, with regard to a knowledge base, if all of its of its defeaters are overruled, being defeated by further justified arguments.[1]

Here I shall not view neither facts nor norms are inference rules of the argumentation system, but rather as premises for it, i.e., as part of its knowledge base (see Prakken and Sartor 2013). Thus, I shall assume a general pattern for building strict inference rules out of modus ponens entailments, and similarly, a general pattern for building defeasible inference rules and undercutters out of defeasible

[1]Argumentation-based semantics (Dung 1995) provides various ways to identify justified arguments, which is done by building maximal sets (called extensions) of the available arguments. For our purposes we can characterise justified arguments as those belonging to an extension that is constructed as follows. We start with the empty set, and progressively admit those arguments that satisfy both of the following conditions: (a) they do not conflict with arguments already admitted, and (b) all their defeaters are defeated by arguments already admitted. The fix-point of this contraction (the set to which no further arguments can be added that satisfy the conditions above) is the so-called grounded extension of an argumentation framework. The same outcome can also be obtained though a dialogue game (Prakken and Sartor 1996; Prakken 2001).

rules (such as, norms). For my purpose, I do not need to address preferences between rules. Therefore the following characterisation of a normative argumentation system will suffice.

Definition 4 (Argumentation system). An argumentation system S is a tuple $N_S = (L, R_s, R_d)$ where

- $\mathscr{L}$ is a logical language (here including, in particular, the constructs for propositional logic, action and deontic logic, and the conditional symbols $\Rightarrow$, and $\not\Rightarrow$).
- R_s is the set of all strict inference rules, including
  - Strict modus ponens inference rules: all rules corresponding to the schema $[\phi, \phi \Rightarrow \psi] \mapsto \psi$ for any ϕ and ψ in $\mathscr{L}$;
  - Specification: all rules corresponding to the schema $[\phi] \mapsto \phi[t]$, where t is a substitution of variables in ϕ with terms in $\mathscr{L}$;
  - Logical axioms: all rules corresponding to the schema $[] \mapsto \phi$, where ϕ is any theorem of propositional logic or other deductive logical systems to be used (here action logic **E**, and standard deontic logic **D**}
- R_d is the set of all defeasible inference rules, including
  - Defeasible modus ponens inference rules: all rules corresponding to the schema $[\phi, \phi \Rightarrow \psi] \rightsquigarrow \psi$ for any ϕ and ψ in $\mathscr{L}$.
  - Defeasible undercutting inference rules: all rules corresponding to the schema $[\phi \not\Rightarrow \psi] \rightsquigarrow \neg$‘$[\phi, \phi \Rightarrow \psi] \rightsquigarrow \psi$’ where ‘$[\phi, \phi \Rightarrow \psi] \rightsquigarrow \psi$’ is the name for the inference rule $[\phi, \phi \Rightarrow \psi] \rightsquigarrow \psi$

We read the name of an inference rule as the assertion that the rule is applicable, and so the negation the name of an inference rule is the assertion that the rule is inapplicable; the defeasible undercutting inference schema says that if a norm fails to support its conclusion, then the inference rule based on that norm is inapplicable.

The logical-axioms inference-schema allows any theorem of the deductive logics being used (e.g. $O\phi \rightarrow P\phi$ from deontic logic) to be introduced in any argument, as an unchallengeable premise. We can now define the idea of a normative knowledge base.

Definition 5 (Normative Knowledge Base). A normative knowledge base K is a tuple (C, N), of two sets of premises:

- a set C of contextual circumstances,
- a set N of norms (a normative system)

By contextual circumstances I mean the propositions describing the relevant facts of the case, such as the fact that *John* damages *Tom*, the amount of the damage, whether *John* intended to cause the damage or was careless, etc. This notion of a fact is a relative one, since certain normative rules (the constitutive ones) may establish under what conditions a certain qualification is satisfied, so that an apparently factual qualification becomes a normative outcome. Consider for instance legal rules establishing what counts as negligence in road traffic, or in medical practice. For our purpose however, we do not need to address this issue, since we view as contextual circumstances all relevant true propositions that are not established through the application of the normative system we are considering (i.e., that are not established as being the consequent of a norm whose antecedent condition is satisfied).

Finally we define the notion of an entailment with regard to a normative knowledge base and a normative argumentation system.

Definition 6 (Defeasible entailment). We shall say that a normative knowledge base $K = (C, N)$ defeasibly entails A,and write $K \mid\!\sim A$, to mean that knowledge base K enables us to construct a justified argument for A, using the inference rules in argumentation system S.

For instance, given knowledge base $K_1 = (\{a\}, \{a \Rightarrow b\})$, we can construct an undefeated (indeed unattached) argument A_3 for b. Thus we may say that $K_1 \mid\!\sim b$.

Let us now consider knowledge base

$$K_2 = (\{a, c, d\}, \{a \Rightarrow b, c \Rightarrow \neg b, d \Rightarrow (c \not\Rightarrow \neg b)\})$$

This knowledge base enables the construction of argument A_3 for b, as above.

K_2 also enables the construction of arguments $A_4 = \{c\}$, $A_5 = \{c \Rightarrow \neg b\}$ and $A_6 = \{A_4, A_5 \rightsquigarrow \neg b\}$, the latter being a rebutting counterargument to A_3.

However K_2 also provides for the construction of arguments $A_6 = \{d\}$, $A_7 = \{d \Rightarrow (c \not\Rightarrow \neg b)\}$, $A_8 = \{A_6, A_7 \rightsquigarrow c \not\Rightarrow \neg b\}$ and $A_9 = \{A_8 \rightsquigarrow \neg`([c, c \Rightarrow \neg b] \rightsquigarrow \neg b)'.\}$ The last arguments undercuts A_5, so that A_3 is freed from its only attacker and is thus justified.

The following example shows how from a norm and an instance of its antecedent we can defeasibly derive an instance of the conditional's consequent.

$$\{\mathsf{E}_{Tom}Damaged(John), \mathsf{E}_x Damaged(y) \Rightarrow O\mathsf{E}_x Compensated(y)\} \mid\!\sim$$
$$O\mathsf{E}_{Tom}Compensated(John)$$

To execute this inference we just need to instantiate the pattern $[\phi, \phi \Rightarrow \psi] \rightsquigarrow \psi$ into the defeasible inference rule

$$[\mathsf{E}_{Tom}Damaged(John), \mathsf{E}_x Damaged(y)$$
$$\Rightarrow O\mathsf{E}_x Compensated(y)] \rightsquigarrow O\mathsf{E}_x Compensated(y)$$

and apply it to the fact and rule above.

3 Relativised Obligations and Permissions

In addressing compliance we have to connect a normative system N and the factual circumstances C relevant to N's application, in the context of a given argumentation system. Here I am only interested in obligations and institutional facts that are generated by norms in N, when applied to facts in C. Thus we can assume that C contains (or entails) all factual literals (atomic propositions or negations of them) which are true in the real or hypothetical situation in which the norms have to be applied, without considering how the truth of such literals can be established. For simplicity's sake we can limit C to the factual literals that are relevant to the application of norms in N, matching literals in the antecedent of a norm in N. When the considered factual circumstances are those that hold in the real world (rather than in a merely possible situation), i.e., they are the truths relevant to the application of N in the case at hand, I shall denote them through the expression T_N.

I will now introduce the notion of a relativised obligation, namely, a way of expressing the fact that an obligation holds with regard to a normative system and a set of circumstances. A relativised obligation sentence does not express a norm, but it expresses an assertion about the implications of norms (normative systems) and circumstances (in the terminology of Alchourrón 1969 and Alchourrón and Bulygin 1971 such assertions are called "normative propositions").

Definition 7 (Relativised sentences and obligations). We say that any sentence ϕ holds relatively to normative system N and circumstances C, and write $[\phi]_{C,N}$ iff $(C,N) \mid\!\sim \phi$

$$[\phi]_{C,N} \stackrel{\text{def}}{=} (C,N) \mid\!\sim \phi$$

Let the expression $\mathscr{A}_j$ cover both $\mathsf{E}_j A$ and $\neg\mathsf{E}_j A$ and let $\overline{\mathscr{A}_j}$ denote the complement of $\mathscr{A}_j$ ($\overline{\mathscr{A}_j}$ stands for $\neg\mathsf{E}_j\phi$ if $\mathscr{A}_j = \mathsf{E}_j\phi$; it stands for $\mathsf{E}_j\phi$ if $\mathscr{A}_j = \neg\mathsf{E}_j\phi$). Then we can say that it is obligatory to do (or not to do) an action relatively to a certain set of circumstances and a normative system, if such circumstances and system entail that the action ought (or ought not) to be done.

$$\mathbb{O}_{C,N}\mathscr{A}_x \stackrel{\text{def}}{=} (C,N) \mid\!\sim O\mathscr{A}_x$$

When we are referring to the true relevant circumstances of the real world, denoted as T_N, rather than to circumstances of hypothetical situations, we simply write $[\phi]_N$, or $\mathbb{O}_N\mathsf{E}_x A$.

$$[\phi]_N \stackrel{\text{def}}{=} (T_N,N) \mid\!\sim \phi$$
$$\mathbb{O}_N\mathscr{A}_x \stackrel{\text{def}}{=} (T_N,N) \mid\!\sim O\mathscr{A}_x$$

For instance, let us consider the following example, where N_1 includes a simplified version of the three norms above, and circumstances C_1 are limited to the fact that *John* injured *Tom*:

Example 1.

$$\begin{aligned} C_1 &= \{\mathsf{E}_{John} Injured(Tom)\} \\ N_1 &= \{\mathsf{E}_x Injured(y) \Rightarrow \mathsf{E}_x Damaged(y) \\ &\quad O\neg\mathsf{E}_x Damaged(y) \\ &\quad \mathsf{E}_x Damaged(y) \Rightarrow O\mathsf{E}_x Compensated(y)\} \end{aligned}$$

It is easy to see that the following inferences holds on the basis of example (1):

$$(C_1, N_1) \mid\!\sim \mathsf{E}_{John} Damaged(Tom)$$
$$(C_1, N_1) \mid\!\sim O\mathsf{E}_{John} Compensated(Tom)$$

Therefore, we can say that *John* has damaged *Tom* and that it is obligatory that *John* compensates *Tom*, relatively to N_1 and C_1, i.e., that

$$[\mathsf{E}_{John} Damaged(Tom)]_{N_1, C_1}$$
$$\mathbb{O}_{N_1, C_1} \mathsf{E}_{John} Compensated(Tom)$$

If *John* has really injured *Tom* (and no other relevant circumstances obtain, such as exceptions excluding the application of the norms at issue) we can simply say that, according to N_1, *John* has damaged *Tom* and it is obligatory that *John* compensates *Tom* i.e.:

$$[\mathsf{E}_{John} Damaged(Tom)]_{N_1} \wedge \mathbb{O}_{N_1} \mathsf{E}_{John} Compensated(Tom)$$

On the basis of example (1) we can also say that it is obligatory that *John* refrains from damaging *Tom*

$$\mathbb{O}_{N_1} \neg\mathsf{E}_{John} Damaged(Tom)$$

Given that it holds that $[\mathsf{E}_{John} Damaged(Tom)]_{N_1}$ we can conclude that the latter obligation has been violated, on the basis of the following definition.

Definition 8 (Violation). An obligation $O\mathsf{E}_x A$ of a normative system N is violated in circumstances C iff $(C, N) \mid\!\sim O\mathsf{E}_x A \wedge \neg\mathsf{E}_x A$, In other words the obligation is violated in C, iff both $\mathbb{O}_{C,N} \mathsf{E}_x A$ and $[\neg\mathsf{E}_x A]_{C,N}$ hold.

Here is another small example. The first norm in N_2 says that if one is in a public place then one is forbidden to smoke. The second says that places open to the public are (count as) public places.

Example 2.

$$C_2 = \{OpenToPublic(LectureRoom), in(John, LectureRoom)\}$$
$$N_2 = \{OpenToPublic(y) \Rightarrow PublicPlace(y)$$
$$PublicPlace(y) \wedge in(x, y) \Rightarrow O\neg \mathsf{E}_x Smoke\}$$

We can say then say that according to N_2 in circumstances C_2 it is obligatory that *John* does not smoke ($\mathbb{O}_{C_2,N_2}\neg\mathsf{E}_{Tom}Smoke$).

Clearly, the language of relativised obligations allows us to say that according to different normative systems different obligations hold. For instance, given that Canon law contains both a universal norm prohibiting the use of contraception and a constitutive rule saying any action meant to make a sex act unfruitful counts as artificial contraception, we can conclude that according to the Canon law a woman, say Ann, is forbidden to take the pill in order to have unfruitful sex acts. Similarly, given that Islamic law contains a norm that prohibits receiving interest on loans of money, we can say that according to Islamic law John is forbidden to receive interest on loans of money.

A notion of relativised permission can be provided that corresponds to the above analysis of obligation. While permissions can be modelled as the negation of prohibitions ($P\mathsf{E}_x A \stackrel{\text{def}}{=} \neg O \neg \mathsf{E}_x A$), relativised permissions can be defined as follows.

Definition 9 (Relativised permission). Let us say that it is permissible relatively to N and C that x does (or not does) an action, iff N and C entail that it is permissible to do (or not to do) that action:

$$\mathbb{P}_{C,N}\mathscr{A}_x \stackrel{\text{def}}{=} (C, N) \mid\!\sim P\mathscr{A}_x$$

Note that according to this definition, saying that an action $\mathsf{E}_x\phi$ is permissible relatively to normative system N and circumstances C ($\mathbb{P}_{C,N}\mathsf{E}_x\phi$) does not amount to saying that it is not the case that $\mathsf{E}_x\phi$ is forbidden relatively to the same system and circumstances ($\neg\mathbb{O}_{C,N}\neg\mathsf{E}_x\phi$). Proposition $\mathbb{P}_{C,N}\mathsf{E}_x\phi$ is not equivalent to $\neg\mathbb{O}_{C,N}\neg\mathsf{E}_x\phi$, since the former holds when (C, N) entails $P\mathsf{E}_x\phi$, while the latter holds when (C, N) does not entail $O\neg\mathsf{E}_x\phi$ (see Alchourrón 1969; Alchourrón and Bulygin 1971).

4 Reasoning with Normative Systems

Let us assume that *Tom* wants to know his position concerning the normative systems L (the law). In particular *Tom* is now wondering whether he should pay income tax on the capital gains he obtained by selling his house. Being committed to comply with the law, but not knowing what the law requires from him, *Tom* asks the tax expert *Ann* for advise. Assume that the *Ann* remembers that there is a rule in the tax code that establishes the requirement to pay income taxes on capital gains, but vaguely remembers that there are exceptions to it. This prompts *Ann* to look for exceptions, and she finds indeed one concerning houses. This exception says (in a simplified form) that capital gains from the sale of houses purchased more than 5 year before the sale and inhabited by the seller are exempted from income tax. Assume that Ann's inquiry has led her to conclude that the legal system L she is considering, for instance Italian law, contains the following relevant norms:

$$\begin{aligned} L \supseteq \{ & SellsHouse(x) \Rightarrow O\mathsf{E}_x PayIncomeTaxOnSale, \\ & BoughtMoreThan5YearsBefore(x) \wedge HasInhabitedHouse(x) \\ & \Rightarrow (SellsHouse(x) \not\Rightarrow O\mathsf{E}_x PayIncomeTaxOnSale)\} \end{aligned} \tag{1}$$

where the second norms in (1) says that under the indicated conditions the first one does not hold (is not applicable).

Ann then asks *Tom* whether, at the time of the sale, more that 5 years had elapsed from the *Tom*'s purchase, and whether he has been living in the house. Assume that *Tom* replies positively to the first question and negatively to the second one. Then *Ann* says: "Dear, *Tom*, unfortunately you are legally bound to pay income tax on your gains". In fact, by combining the Italian law L with these factual circumstances (let us assume these circumstances are the only relevant ones), *Ann* can see that the following inference holds:

$$(\{SellsHouse(Tom), \neg HasInhabitedHouse(Tom)\}, L) \mathrel{|\!\sim} O\mathsf{E}_{Tom} PayIncomeTaxOnSale$$

so that, given that both factual premises are true, she can infer what she tells her client:

$$\mathbb{O}_L \mathsf{E}_{Tom} PayIncomeTaxOnSale$$

If *Tom* asks for an explanation, *Ann* would probably answer by saying that whenever one has not lived in the house one sells, then according to the law one has the obligation to pay income tax:

$$SellsHouse(x) \wedge \neg HasInhabitedSoldHouse(x) \Rightarrow \mathbb{O}_L \mathsf{E}_{(}x) PayIncomeTaxOnSale \tag{2}$$

Note that formula (2) does not express a norm of L (there is no norm in L which has exactly that content, see formula (1)). More generally (2) is no norm at all, but rather is a general conditional statement about L, namely the statement that in case that the seller has not inhabited the sold house, then L entails that the seller has to pay taxes on capital gains. Similarly, if *Ann* were contacted by *Tom* before making the sale, she would tell him: "Since you have not inhabited the house, if you seel it you will have to pay income tax on your capital gain".

5 Dynamic Normative Systems

Let us now consider how an agent (a legislator) can have the ability to introduce new norms in N. For this purpose, we need to assume that N is a dynamic normative system (Kelsen 1967), including meta-norms which determine what new norms will be valid according to N.

In the framework we have described above, the idea of such a metanorm can be captured through an additional pattern for defeasible inference rules, which enables the production basic norm set to be expanded by further norms. Thus we obtain what we may call a dynamic extension of the normative argumentation system.

Definition 10 (Dynamic Normative Argumentation System). A dynamic normative argumentation systems DS is obtained by adding to an argumentation systems S the following inference rules

- Norm creation: all inference rules corresponding to the schema $Valid(\phi) \mapsto \phi$, for any norm ϕ in $\mathscr{L}$.

Note that I prefer to model this principle as a strict rule, but depending on how we understand the notion of validity, we could also model it as a defeasible rule (see Sartor 2008).

In a dynamic normative argumentation system, arguments may use rules that do not belong to an initial knowledge base, but that are qualified as valid by norms in that knowledge base. So, let us assume that the knowledge base K of a normative system includes a meta-norm saying that whatever norm ϕ is issued by the legislator Leg than ϕ is valid (for simplicity's sake I do not consider the temporal dimension of validity, see Governatori and Rotolo 2010).

$$\mathsf{E}_{Leg}Issued(\phi) \Rightarrow Valid(\phi)$$

Given this background, let us assume that legislator accomplished the action of issuing a new norm, for instance, a norm prohibiting any agent x to smoke:

$$\mathsf{E}_{Leg}Issued(O\neg\mathsf{E}_x Smoke)$$

The accomplishment of the action described in this formula is a new fact, which is added to the true factual circumstances T_N. Thus we can build the following sequence of arguments

- $A_1 = \mathsf{E}_{Leg}Issued(O\neg\mathsf{E}_x Smoke)$, premise;
- $A_2 = \mathsf{E}_{Leg}Issued(\phi) \Rightarrow Valid(\phi)$, premise;
- $A_3 = A_2 \mapsto (\mathsf{E}_{Leg}Issued(O\neg\mathsf{E}_x Smoke) \Rightarrow Valid(O\neg\mathsf{E}_x Smoke))$, by specification;
- $A_4 = A_1, A_3 \rightsquigarrow Valid(O\neg\mathsf{E}_x Smoke)$, by defeasible modus ponens;
- $A_5 = A4 \mapsto O\neg\mathsf{E}_x Smoke$, by norm creation;
- $A_6 = A5 \mapsto O\neg\mathsf{E}_{Tom} Smoke$, by specification.

Thus, on the basis of argument A_6 (which we assume to be unchallenged) we can conclude that smoking is forbidden to *Tom* according to *N*:

$$\mathbb{O}_N \neg \mathsf{E}_{Tom} Smoke$$

6 Conclusion

In this paper I have shown how a reasoner may approach a normative system, namely a distinct set of norms, viewing it as an object that enables the derivation of normative conclusions that are relative to that system. For this purpose I have first considered how to model actions, obligations and norms. Then I have defined an argumentation system which takes as inputs knowledge bases of facts and norms, and produces appropriate arguments. On this basis I have considered how obligations and permission can be relative to a particular normative systems, and I have provided a meta-logical representation of this idea. Finally I have developed some considerations on how to model dynamic normative systems in this framework.

While this work is still very preliminary, I hope it can provide some clues on how to model metalevel reasoning with normative systems. Obvious extensions, to be considered in future work, concern integrating this idea with the decision-making process of the concerned agents (for a preliminary attempt, see Sartor 2012), and modelling reasoning with multiple distinct normative systems.

References

Alchourrón, C.E. 1969. Logic of norms and logic of normative propositions. *Logique et analyse* 12: 242–268.

Alchourrón, C.E., and E. Bulygin. 1971. *Normative systems*. Vienna: Springer.

Belnap, N., M. Perloff, M. Xu, and P. Bartha. 2001. *Facing the future: Agents and choices in our indeterminist world*. Oxford: Oxford University Press.

Dung, P.M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*–person games. *Artificial Intelligence* 77: 321–357.

Governatori, G., and A. Rotolo. 2010. Changing legal systems: Legal abrogations and annulments in defeasible logic. *Logic Journal of IGPL* 18: 157–194.

Jones, A.J., and M.J. Sergot. 1996. A formal characterisation of institutionalised power. *Journal of the IGPL* 4: 429–445.

Kelsen, H. 1967. *The pure theory of law*. Berkeley: University of California Press.

Pörn, I. 1977. *Action theory and social science: Some formal models*. Dordrecht: Reidel.

Prakken, H. 2001. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese* 127: 187–219.

Prakken, H. 2010. An abstract framework for argumentation with structured arguments. *Argument and Computation* 1: 93–124.

Prakken, H., and G. Sartor. 1996. Rules about rules: Assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law* 4: 331–368.

Prakken, H., and G. Sartor. 2013. Formalising arguments about norms. In *Proceeding of JURIX 2013: The twenty-sixth annual conference on legal knowledge and information systems*, 121–30. Amsterdam: IOS.

Sartor, G. 2008. Legal validity: An inferential analysis. *Ratio Juris* 21: 212–247.

Sartor, G. 2012. Intentional compliance with normative systems. In *The goals of cognition. Essays in honor of Cristiano Castelfranchi*, ed. F. Paglieri, L. Tummolini, R. Falcone, and M. Miceli, 627–656. London: College Publications.

Searle, J.R. 1995. *The construction of social reality*. New York: Free.

Sergot, M.J. 2001. A computational theory of normative positions. *ACM Transactions on Computational Logic* 2: 581–662.

van der Torre, L., and J. Hansen. 2008. In *Deontic logic in computer science: Course material. 20th European summer school in logic, language and information*, Hamburg.