Logic and **L**ogical **P**hilosophy
Volume 32 (2023), 583–617
DOI: 10.12775/LLP.2023.004

Piotr Kulicki, Robert Trypuz
Robert Craven, Marek Sergot

A Unified Logical Framework for Reasoning about Deontic Properties of Actions and States

Abstract. This paper studies some normative relations that hold between actions, their preconditions and their effects, with particular attention to connecting what are often called 'ought to be' norms with 'ought to do' norms. We use a formal model based on a form of transition system called a 'coloured labelled transition system' (coloured LTS) introduced in a series of papers by Sergot and Craven. Those works have variously presented a formalism (an 'action language') $n\mathcal{C}+$ for defining and computing with a (coloured) LTS, and another, separate formalism, a modal language interpreted on a (coloured) LTS used to express its properties. We consolidate these two strands. Instead of specifying the obligatory and prohibited states and transitions as part of the construction of a coloured LTS as in $n\mathcal{C}+$, we represent norms in the modal language and use those to construct a coloured LTS from a given regular (uncoloured) one. We also show how connections between norms on states and norms on transitions previously treated as fixed constraints of a coloured LTS can instead be defined within the modal language used for representing norms.

Keywords: deontic logic; ought to do; ought to be; $n\mathcal{C}+$; deontic action logic; transition system; conditional norms

1. Introduction

Normative regulations consist of norms having different perspectives on agents' behaviour. Some norms refer explicitly to actions, and the names of desired, undesired or permitted actions appear in them. Other norms refer to states of affairs resulting from agents' activities. In this case

Received July 4, 2022. *Accepted* April 18, 2023. *Published online* June 2, 2023

it is the states of affairs that are the subjects of norms rather than the specific actions that bring them about.

For example, in academic life we know that for the positive assessment of a researcher a certain number of papers should be published within the period under assessment (so there might be a norm that X papers should have been published in prestigious journals in the previous Y years) or that (in Poland) an assistant professor nine years after appointment should have the 'habilitation' degree or that (in the UK) the primary supervisor of a doctoral student must be the holder of a permanent academic position. Those are norms on states; the specific acts required to achieve them are not prescribed. On the other hand, universities also establish norms on actions, regulating *how* agents should *act*, e.g., that teachers must not disclose a student's personal file to any third party, that professors must conduct a review of their research assistants annually on the anniversary of the appointment, or that students are not permitted to use printers in staff offices. And of course then there are (norms on acts) no smoking, no copying, no fighting, no opening of fire doors except in an emergency, ..., as well as (norms on states) no food in laboratories, no mobile telephones in examination rooms, no more than four persons in a lift, and so on.

Traditionally this difference is presented as the distinction between *ought to do* and *ought to be* norms. In many normative systems, perhaps even all, both kinds of norms are present at the same time. Even in criminal law, which is often said to be concerned exclusively with acts, it is generally recognised that this is an over-simplification: it may be prohibited to be in possession of a certain drug, or to be carrying a weapon in public, or to be on certain kinds of private property without authorisation. It is difficult to insist that these are examples of ought to do norms, and natural to see them as instances of ought to be. A formal account of normative systems requires that both kinds of norms can be adequately expressed.

To study the normative relations between actions and states we use a formal model in the form of a labelled transition system (LTS). This structure makes it possible to consider properties of both states of affairs and transitions between them, and the representation of actions and action types.

More precisely, we shall base our work on the 'coloured transition systems' and the $n\mathcal{C}+$ formalism presented in a series of papers by Sergot and Craven [3, 9, 10, 11, 12]. Coloured transition systems partition the states and transitions of an LTS into those that are acceptable/permitted ('green') and those that are not ('red'). $n\mathcal{C}+$ is a formalism for defining and computing with a coloured LTS. It is itself an extension of the action language $\mathcal{C}+$ [4] — a formalism for specifying and reasoning about the effects of actions and the persistence of facts over time.

The problem of relations between norms on states and norms on actions was recently discussed by some of us in [7]. The coloured LTS framework enables us to express the ideas of connecting norms on states with norms on actions more easily than in the approach presented there.

$n\mathcal{C}+$ is a mature formalism with quite extensive theory exploring its expressive power and formal properties, computer implementations, and experience with applications to examples. Agency was introduced in [3, 10]. Examples showing how it may be applied to scenarios typical for AI research are available [10, 11].

The emphasis in those previous works varies. In some [3, 9, 12] the focus is on presenting $n\mathcal{C}+$ as a formalism (an 'action language') for *defining* and computing with a (coloured) LTS. The language used to express properties of a (coloured) LTS was secondary, and was essentially the query language used with the CCalc implementation of $\mathcal{C}+$. In other works [10, 11], the action language $n\mathcal{C}+$ does not appear; examples are represented as a (coloured) LTS without discussion of how it is constructed. Instead, a two-sorted multi-modal language is introduced in order to express properties of a coloured LTS. There are thus two separate and quite different formalisms: $n\mathcal{C}+$ which is used to *define* a coloured LTS, and an (unnamed) modal language interpreted on a coloured LTS used to express its properties. In this paper we want to consolidate these two strands. In $n\mathcal{C}+$ one writes rules to specify which states and transitions are to be green/red as part of the definition of a coloured LTS. Here, we will instead represent norms in the modal language and then show how they can be used to construct a coloured LTS from a given regular (uncoloured) one.

In coloured LTS the main connection between norms on states and norms on actions is a well-formedness constraint incorporated as a fixed feature of every coloured LTS. That constraint — the 'green-green-green' constraint — requires that 'a green (permitted, acceptable, legal) transition starting from a green (permitted, acceptable, legal) state must always lead to a green (acceptable, legal, permitted) state'. A coloured LTS that fails to satisfy this constraint is not well defined. The second contribution of the paper is to treat this constraint differently. Instead of requiring it to be a fixed feature of every coloured LTS it is defined within the modal language used for representing norms. It thereby becomes optional (though generally desirable). Other possible variations can be treated similarly.

Previous works on $n\mathcal{C}+$ and coloured LTS have dealt with the multi-agent case, agent-specific norms and the incorporation of a logic of agency. Those topics are peripheral to the main concerns of this paper and will be omitted.

The paper is organised as follows. Section 2 presents LTS and the modal language used for expressing their properties, Section 3 presents coloured LTS, Section 4 shows how unconditional norms of prohibition and obligation, both on actions and states of affairs, are defined, and Section 5 discusses the representation of conditional norms. Section 6 summarises and adds some technical detail.

2. Preliminaries: Transition systems

A *labelled transition system* (LTS) is a structure of the form

$$\langle S, A, R, \mathit{prev}, \mathit{post}, \mathit{label} \rangle$$

where S is a (non-empty) set of *states*, A is a (non-empty) set of *action types*, also called '*labels*', and R is a (non-empty) set of *transitions*. *prev* and *post* are functions from R to S: $\mathit{prev}(\tau)$ denotes the initial state of a transition τ, and $\mathit{post}(\tau)$ its resulting (or end) state. *label* is a function from R to A: $\mathit{label}(\tau)$ denotes the action type (or event type depending on context) performed in transition τ.

For the common special case of a LTS in which the transitions are triples $R \subseteq S \times A \times S$, $\mathit{prev}((s, \varepsilon, s')) = s$, $\mathit{post}((s, \varepsilon, s')) = s'$, and $\mathit{label}((s, \varepsilon, s')) = \varepsilon$. All the examples in this paper are of this special form.

The account can be generalised to deal with the multi-agent case [see, e.g., 3, 10]. *label* is replaced by a function $\mathit{strand}\colon Ag \times R \to A$ where Ag is a (finite) set of agent names and $\mathit{strand}(x, \tau)$ picks out the action type performed by agent x in transition τ. We will not deal with the multi-agent case in this paper; the methods work in the same way.

A transition system may be indeterministic. There may be several different transitions from the same initial state and with the same label but with different resulting states. There may also be several different transitions with the same initial and final states but with different labels, representing different ways of reaching the same result. A transition in LTS is more than just an ordered pair connecting two states.

Language. Given a labelled transition system, it is usual to define a language of propositional atoms or 'state variables' in order to express properties of states. We employ a *two-sorted* language. We have a set P_S of propositional atoms for expressing properties of states, and a disjoint set P_R of propositional atoms for expressing properties of transitions. Models are structures

$$\mathcal{M} = \langle S, A, R, \mathit{prev}, \mathit{post}, \mathit{label}, h_S, h_R \rangle$$

where $h_S \colon P_S \to 2^S$ and $h_R \colon P_R \to 2^R$ are the valuation functions for P_S in states S and P_R in transitions R, respectively. In this paper, unless stated otherwise, the atoms P_R are used to represent properties of the action(s) performed in a transition.

Formulas are *state formulas* and *transition formulas*. State formulas are:

$$F \ ::= \ \text{any atom } p \text{ of } P_S \mid \neg F \mid F \wedge F \mid \Box F \mid \underset{\rightarrow}{\Box} \varphi$$

where φ is any transition formula. Transition formulas are

$$\varphi \ ::= \ \text{any atom } \alpha \text{ of } P_R \mid \neg \varphi \mid \varphi \wedge \varphi \mid 0{:}F \mid 1{:}F \mid \ \boxdot \varphi$$

where F is any state formula. We have the usual truth-functional abbreviations. $\Diamond$ is the dual of $\Box$ and $\underset{\rightarrow}{\Diamond}$ is the dual of $\underset{\rightarrow}{\Box}$.

Semantics. Truth-functional connectives have the usual interpretations for both state and transition formulas. The satisfaction definitions for the other operators are as follows.

State formulas:

$$\mathcal{M}, s \models \Box F \quad \text{iff} \quad \mathcal{M}, s' \models F \text{ for every } s' \in S$$
$$\mathcal{M}, s \models \underset{\rightarrow}{\Box} \varphi \quad \text{iff} \quad \mathcal{M}, \tau \models \varphi \text{ for every } \tau \in R \text{ such that } \mathit{prev}(\tau) = s$$

$\Box$ is the S5 universal necessity for states. $\underset{\rightarrow}{\Box} \varphi$ is true at a state s when every transition from state s satisfies φ. $\underset{\rightarrow}{\Diamond} \varphi$ says that there is a transition of type φ from the current state.

Transition formulas:

$$\begin{aligned}
\mathcal{M}, \tau \models\ & 0{:}F \ \text{ iff } \ \mathcal{M}, \mathit{prev}(\tau) \models F \\
\mathcal{M}, \tau \models\ & 1{:}F \ \text{ iff } \ \mathcal{M}, \mathit{post}(\tau) \models F \\
\mathcal{M}, \tau \models\ & \boxdot\varphi \ \text{ iff } \ \mathcal{M}, \tau' \models \varphi \text{ for every } \tau' \text{ such that } \mathit{prev}(\tau) = \mathit{prev}(\tau')
\end{aligned}$$

A transition is of type $0{:}F$ when its initial state satisfies the state formula F, and of type $1{:}F$ when its resulting state satisfies F. It is of type $\boxdot\varphi$ when all transitions from the same initial state are of type φ. (The $\boxdot$ modality is not used in this paper. It is an essential component when the language is extended with agency operators as in [10, 11]. It is left here for compatibility with these other papers.)

As usual, we say a state formula F is *valid* in a model $\mathcal{M}$, written $\mathcal{M} \models F$, when $\mathcal{M}, s \models F$ for every state s in S, and a transition formula φ is *valid* in a model $\mathcal{M}$, written $\mathcal{M} \models \varphi$, when $\mathcal{M}, \tau \models \varphi$ for every transition τ in R. A formula is *valid* if it is valid in every model (written $\models F$ and $\models \varphi$, respectively). We will write $\varphi \equiv \psi$ when $\varphi \leftrightarrow \psi$ is valid, and $\varphi \equiv_{\mathcal{M}} \psi$ when $\varphi \leftrightarrow \psi$ is valid in a model $\mathcal{M}$, and likewise for state formulas.

We use the following notation for 'truth sets':

$$\|F\|^{\mathcal{M}} =_{\text{def}} \{s \in S \mid \mathcal{M}, s \models F\}; \quad \|\varphi\|^{\mathcal{M}} =_{\text{def}} \{\tau \in R \mid \mathcal{M}, \tau \models \varphi\}.$$

Remarks. The operators 0: and 1: are not normal in the usual sense because formulas F and $0{:}F$ (and $1{:}F$) are of different sorts. However, they behave like normal operators in the sense that, for all $n \geq 0$, if $F_1 \wedge \cdots \wedge F_n \rightarrow F$ is valid then so are $0{:}F_1 \wedge \cdots \wedge 0{:}F_n \rightarrow 0{:}F$ and $1{:}F_1 \wedge \cdots \wedge 1{:}F_n \rightarrow 1{:}F$.

Since *prev* and *post* are (total) functions on R, we have

$$0{:}F \equiv \neg 0{:}\neg F \qquad \text{and} \qquad 1{:}F \equiv \neg 1{:}\neg F$$

(and 0: and 1: distribute over all truth-functional connectives).

The state formula $\underset{\rightarrow}{\Diamond}\varphi$ says that there is a transition of type φ from the current state, or in the terminology of transition systems, that φ is 'executable'. $\underset{\rightarrow}{\Diamond} 1{:}F$ expresses that there is a transition from the current state to a state where F is true. $\underset{\rightarrow}{\Box}(\varphi \rightarrow 1{:}F)$ says that all transitions of type φ from the current state result in a state where F is true.

$\underset{\rightarrow}{\Box}$ is normal in the sense that, for all $n \geq 0$, if $\varphi_1 \wedge \cdots \wedge \varphi_n \rightarrow \varphi$ is valid then so is $\underset{\rightarrow}{\Box}\varphi_1 \wedge \cdots \wedge \underset{\rightarrow}{\Box}\varphi_n \rightarrow \underset{\rightarrow}{\Box}\varphi$.

The relationship between state formulas and transition formulas is easily established. The following are valid. The first is a transition formula. The second is a state formula.

$$\boxdot\varphi \leftrightarrow 0{:}\underset{\rightarrow}{\Box}\varphi$$
$$F \rightarrow \underset{\rightarrow}{\Box}0{:}F \tag{1}$$

Completeness is easy to show, e.g., via canonical models.

If we further assume that the LTS is serial, i.e., that for every s in S there is τ in R such that $prev(\tau) = s$, then $\underset{\rightarrow}{\Diamond}\top$ is valid and axiom schema (1) is strengthened to $\underset{\rightarrow}{\Box}0{:}F \leftrightarrow F$. Without seriality we have only the weaker $\underset{\rightarrow}{\Diamond}\top \rightarrow (\underset{\rightarrow}{\Box}0{:}F \rightarrow F)$, which is already implied by schema (1).

Generally speaking, properties of labelled transition systems can be expressed either as transition formulas or as state formulas. For this paper state formulas are more convenient. Clearly $\mathcal{M} \models \varphi$ (a transition formula) iff $\mathcal{M} \models \underset{\rightarrow}{\Box}\varphi$ (a state formula). And $\mathcal{M} \models F$ (a state formula) implies $\mathcal{M} \models 0{:}F \wedge 1{:}F$ (a transition formula); the implication is an equivalence if the LTS is serial (i.e., when $\mathcal{M} \models \underset{\rightarrow}{\Diamond}\top$). This relationship can be expressed using the universal modality $\Box$ for states, as the validity of:

$$\Box F \rightarrow \underset{\rightarrow}{\Box}(0{:}F \wedge 1{:}F) \tag{2}$$

For future reference in the discussion of conditional norms (Section 5), note that the following property is derivable from axiom schema (1):

$$(F \rightarrow \underset{\rightarrow}{\Box}\varphi) \equiv \underset{\rightarrow}{\Box}(0{:}F \rightarrow \varphi) \tag{3}$$

PROOF. Right-to-left: $\models \underset{\rightarrow}{\Box}(0{:}F \rightarrow \varphi) \rightarrow (\underset{\rightarrow}{\Box}0{:}F \rightarrow \underset{\rightarrow}{\Box}\varphi)$. And $\models F \rightarrow \underset{\rightarrow}{\Box}0{:}F$ (axiom (1)). Hence $\models \underset{\rightarrow}{\Box}(0{:}F \rightarrow \varphi) \rightarrow (F \rightarrow \underset{\rightarrow}{\Box}\varphi)$.

Left-to-right (contrapositive): $\models \neg\underset{\rightarrow}{\Box}(0{:}F \rightarrow \varphi) \rightarrow \underset{\rightarrow}{\Diamond}(0{:}F \wedge \neg\varphi)$ and $\models \underset{\rightarrow}{\Diamond}(0{:}F \wedge \neg\varphi) \rightarrow (\underset{\rightarrow}{\Diamond}0{:}F \wedge \underset{\rightarrow}{\Diamond}\neg\varphi)$. $\models \underset{\rightarrow}{\Diamond}0{:}F \rightarrow F$ (axiom (1), dual). Hence $\models (\underset{\rightarrow}{\Diamond}0{:}F \wedge \underset{\rightarrow}{\Diamond}\neg\varphi) \rightarrow (F \wedge \neg\underset{\rightarrow}{\Box}\varphi)$. And $(F \wedge \neg\underset{\rightarrow}{\Box}\varphi) \equiv \neg(F \rightarrow \underset{\rightarrow}{\Box}\varphi)$. ⊣

3. Coloured transition systems

A *coloured transition system* is a labelled transition system in which the states S and transitions R are both classified into two (possibly empty) categories:

- $S_g \subseteq S$ is the set of 'permitted' ('acceptable', 'legal') states — we call S_g the 'green' states; $S_{red} = S \setminus S_g$ are the 'red states'.
- $R_g \subseteq R$ is the set of 'permitted' ('acceptable', ' 'legal') transitions — we call R_g the 'green' transitions; $R_{red} = R \setminus R_g$ are the 'red transitions'.

Semantical devices which partition states (and here, transitions) into two categories are common in the field of deontic logic [see, e.g., 1, 5].

[12] presents a refinement in which the states of a transition systems are ordered depending on how well each complies with a set of explicitly stated norms. We will stick to a simple binary classification in this paper.

We would get more precision by colouring *paths/runs* of the transition system (i.e., sequences of transitions) instead of just single transitions. One could then extend the logics presented in this paper with features from a temporal logic such as CTL. We will not do that here.

Instead of introducing a special category of coloured transition systems with extra components S_g and R_g as in [3, 12], we now prefer to introduce colourings for states and transitions by means of suitably chosen propositional atoms. In particular, let the state atom red_s represent that a state is red, and the transition atom red_t that a transition is red. Let $green_s$ and $green_t$ be abbreviations for $\neg red_s$ and $\neg red_t$, respectively. A coloured LTS is then a structure of the form

$$\mathcal{M} = \langle \mathcal{T}, h^n_S, h^n_R \rangle$$

where $\mathcal{T} = \langle S, A, R, prev, post, label, h_S, h_R \rangle$ is a labelled transition system and where h^n_S and h^n_R are the valuation functions for the distinguished atoms red_s and $green_s$ in states S and red_t and $green_t$ in transitions R, respectively. $h^n_S(red_s) = \|red_s\|^{\mathcal{M}}$ denotes the 'red states' of $\mathcal{M}$ and $\|green_s\|^{\mathcal{M}} = S \setminus \|red_s\|^{\mathcal{M}}$ its 'green states'; $h^n_R(red_t) = \|red_t\|^{\mathcal{M}}$ denotes the 'red transitions' and $\|green_t\|^{\mathcal{M}} = R \setminus \|red_t\|^{\mathcal{M}}$ the 'green transitions'. We say that $\mathcal{M}$ is a *colouring* of the LTS $\mathcal{T}$. We sometimes write $\mathcal{M} = \langle \mathcal{T}, S_{red}, R_{red} \rangle$ where $S_{red} = h^n_S(red_s)$ and $R_{red} = h^n_R(red_t)$.

In previous works, a coloured transition system was further required to satisfy a kind of *well-formedness* principle: a green (permitted, acceptable, legal) transition in a green (permitted, acceptable, legal) state must always lead to a green (acceptable, legal, permitted) state. This is referred to as the *green-green-green* constraint, or *ggg* for short. (A similar constraint for a different formal framework is found in [2].) A coloured LTS which fails to satisfy this constraint is not well defined.

The *ggg* constraint can be expressed as validity in the coloured LTS of the state formula

$$green_s \to \underset{\to}{\Box}(green_t \to 1{:}green_s)$$

equivalently

$$green_s \to \underset{\to}{\Box}(1{:}red_s \to red_t)$$

or, equivalently again, of the transition formula

$$(0{:}green_s \wedge 1{:}red_s) \to red_t$$

In the multi-agent case [3, 10, 11] it is often necessary to distinguish between impersonal ('system') norms/colourings and agent-specific norms/colourings. A transition may be unacceptable (illegal, 'red') as regards one particular agent x but not as regards another agent y. In addition to the impersonal/'system' colourings red_t and $green_t$, we then have transition atoms $red_t(x)$ and $green_t(x)$ for every individual agent x. And similarly for states. A state in which an agent x has an unpaid debt may be regarded as bad as regards x, represented $red_s(x)$, but not bad as regards a different agent y. We will not deal with the multi-agent case in this paper. Relationships between impersonal/'system' and agent-specific norms are discussed in the works cited above. The specification of agent-specific colourings/norms works in exactly the same way as for the colourings/norms treated in this paper.

Defined deontic operators. Let state formulas $\mathsf{F}^s F$ and $\mathsf{O}^s F$ represent that a state of affairs where F holds is forbidden/obligatory respectively.

The satisfaction conditions are:

$$\mathcal{M}, s \models \mathsf{F}^s F \quad \text{iff} \quad \|F\|^{\mathcal{M}} \subseteq \|red_s\|^{\mathcal{M}}$$
$$\mathcal{M}, s \models \mathsf{O}^s F \quad \text{iff} \quad \|green_s\|^{\mathcal{M}} \subseteq \|F\|^{\mathcal{M}}$$

We define:

$$\mathsf{F}^s F =_{\text{def}} \Box(F \to red_s)$$
$$\mathsf{O}^s F =_{\text{def}} \Box(green_s \to F)$$

Naturally: $\mathsf{O}^s F \equiv \mathsf{F}^s \neg F$. And $\mathsf{O}^s F \equiv (\neg F \to red_s)$. A state of affairs where F holds is permitted when it is not forbidden, i.e., when $\Diamond(F \wedge green_s)$ holds.

Note that these satisfaction conditions are independent of the state s. These notions of obligation and prohibition are global properties of a *model*.

The obligatory/forbidden transitions from a state are defined in terms of red_t and green_t. Let $\underrightarrow{\mathsf{O}}\,\varphi$ and $\underrightarrow{\mathsf{F}}\,\varphi$ represent that a transition of type φ is obligatory/forbidden respectively in the current state. $\underrightarrow{\mathsf{O}}\,\varphi$ and $\underrightarrow{\mathsf{F}}\,\varphi$ are state formulas.

Let $\mathit{exe}(s)$ denote the set of transitions that have s as their initial state:

$$\mathit{exe}(s) \;=_{\text{def}}\; \{\tau \in R \mid \mathit{prev}(\tau) = s\}$$

The satisfaction conditions are, in terms of $\|\mathit{green}_t\|^{\mathcal{M}}$ and $\|\mathit{red}_t\|^{\mathcal{M}}$:

$$\mathcal{M}, s \models \underrightarrow{\mathsf{O}}\,\varphi \quad \text{iff} \quad \mathit{exe}(s) \cap \|\mathit{green}_t\|^{\mathcal{M}} \subseteq \|\varphi\|^{\mathcal{M}}$$
$$\mathcal{M}, s \models \underrightarrow{\mathsf{F}}\,\varphi \quad \text{iff} \quad \mathit{exe}(s) \cap \|\varphi\|^{\mathcal{M}} \subseteq \|\mathit{red}_t\|^{\mathcal{M}}$$

And so we define state formulas:

$$\underrightarrow{\mathsf{O}}\,\varphi \;=_{\text{def}}\; \underrightarrow{\Box}\,(\mathit{green}_t \to \varphi)$$
$$\underrightarrow{\mathsf{F}}\,\varphi \;=_{\text{def}}\; \underrightarrow{\Box}\,(\varphi \to \mathit{red}_t)$$

The satisfaction conditions for these expressions are not independent of the state.

Again: $\underrightarrow{\mathsf{O}}\,\varphi \equiv \underrightarrow{\mathsf{F}}\,\neg\varphi$. And $\underrightarrow{\mathsf{O}}\,\varphi \equiv \underrightarrow{\Box}\,(\neg\varphi \to \mathit{red}_t)$. Permission is the dual of obligation. A transition of type φ is permitted at state s when $\neg\,\underrightarrow{\mathsf{O}}\,\neg\varphi$ is true at s, i.e., when $\underrightarrow{\Diamond}\,(\varphi \wedge \mathit{green}_t)$ is true at s.

These definitions are essentially instances of the familiar Anderson-Kanger reduction in deontic logic. (See e.g., [1, 5] for a survey and discussion of relationships to Standard Deontic Logic (SDL).) The slight difference here is that we are employing a two-sorted language. That affects $\underrightarrow{\mathsf{O}}$ and $\underrightarrow{\mathsf{F}}$ in particular.

Both prohibition operators enjoy the following properties. These expressions are valid:

$$\underrightarrow{\mathsf{F}}\,(\varphi \vee \psi) \leftrightarrow (\underrightarrow{\mathsf{F}}\,\varphi \wedge \underrightarrow{\mathsf{F}}\,\psi)$$
$$\underrightarrow{\mathsf{F}}\,(\varphi \wedge \neg\varphi)$$
$$\mathsf{F}^s(F \vee G) \leftrightarrow (\mathsf{F}^s F \wedge \mathsf{F}^s G)$$
$$\mathsf{F}^s(G \wedge \neg G)$$

The prohibition operators are cumulative — if two actions (states) are forbidden, then their alternative (being a more general description) is also forbidden — and homogeneous — any realisation of a forbidden action (or any instance of a forbidden state) is forbidden.

In terms of the obligation operators:

$$\underrightarrow{\mathsf{O}}(\varphi \wedge \psi) \leftrightarrow (\underrightarrow{\mathsf{O}}\varphi \wedge \underrightarrow{\mathsf{O}}\psi)$$
$$\underrightarrow{\mathsf{O}}(\varphi \vee \neg\varphi) \tag{4}$$
$$\mathsf{O}^s(F \wedge G) \leftrightarrow (\mathsf{O}^s F \wedge \mathsf{O}^s G)$$
$$\mathsf{O}^s(G \vee \neg G) \tag{5}$$

Note that $\underrightarrow{\mathsf{F}}(\varphi \wedge \neg\varphi)$ is equivalent to $\underrightarrow{\mathsf{O}}(\varphi \vee \neg\varphi)$ and to $\underrightarrow{\mathsf{O}}\top$. $\mathsf{F}^s(G \wedge \neg G)$ is equivalent to $\mathsf{O}^s\top$. These are properties of every normal modal operator. They should not be confused with the property associated with obligation in SDL, which would correspond to the validity of $\neg\underrightarrow{\mathsf{O}}\bot$ and $\neg\mathsf{O}^s\bot$. $\neg\underrightarrow{\mathsf{O}}\bot$ is valid in models which have a seriality property, viz. that from every state there is at least one green transition, i.e., in models where $\underrightarrow{\Diamond}\, green_t$ is valid. $\neg\mathsf{O}^s\bot$ is valid when there is at least one green state, i.e., in models where $\Diamond\, green_s$ is valid. The following properties are valid:

$$(\underrightarrow{\mathsf{O}}\varphi \wedge \underrightarrow{\mathsf{O}}\neg\varphi) \leftrightarrow \underrightarrow{\Box}\, red_t$$
$$(\mathsf{O}^s F \wedge \mathsf{O}^s\neg F) \leftrightarrow \Box\, red_s$$

Properties (4) and (5) are consequences of the more general validities:

$$\underrightarrow{\Box}\varphi \rightarrow \underrightarrow{\mathsf{O}}\varphi$$
$$\Box\varphi \rightarrow \mathsf{O}^s\varphi$$

4. Unconditional norms on actions and states

4.1. Example

Let us start with an example taken from [7].

Example 1. An agent signs a contract to carry out construction works. In the contract there is a description of the desired product, i.e., a state of affairs which is to be attained. At the same time the work activities are regulated, among other things, by safety standards which limit possible actions leading to the desired effects to those which are safe. For instance during the construction works: it is obligatory to use designated passages when moving from one place to another, it is forbidden to throw objects, etc. ⊣

The example can be modelled as an LTS $\mathcal{T}_1$ as depicted in Figure 1.

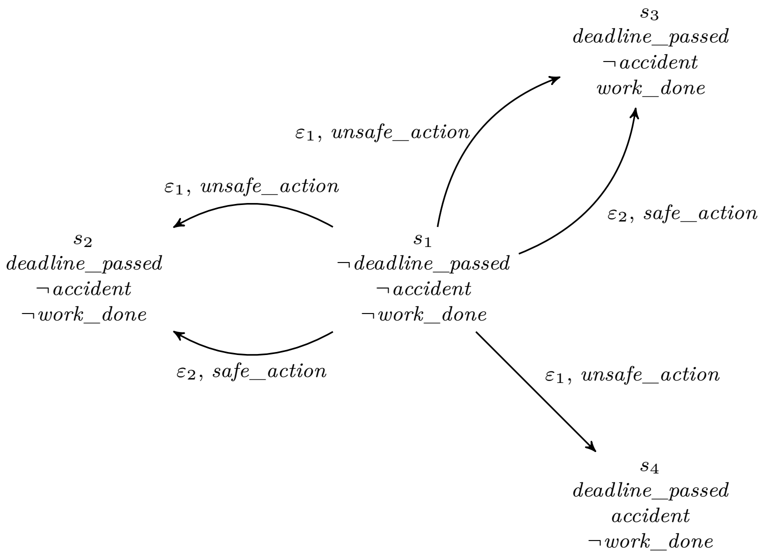

Figure 1. An LTS representation $\mathcal{T}_1$ of Example 1

There are four states in the LTS: s_1, s_2, s_3 and s_4. Each is uniquely described by the combination of three propositions:

- $\|\neg\, deadline\_passed\|^{\mathcal{T}_1} = \{s_1\}$
- $\|deadline\_passed\|^{\mathcal{T}_1} = \{s_2, s_3, s_4\}$
- $\|accident\|^{\mathcal{T}_1} = \{s_4\}$
- $\|work\_done\|^{\mathcal{T}_1} = \{s_3\}$

The figure shows the transitions from state s_1. There are five possible transitions of two types:

- $\|safe\_action\|^{\mathcal{T}_1} = \{(s_1, \varepsilon_2, s_2), (s_1, \varepsilon_2, s_3)\}$
- $\|unsafe\_action\|^{\mathcal{T}_1} = \{(s_1, \varepsilon_1, s_2), (s_1, \varepsilon_1, s_3), (s_1, \varepsilon_1, s_4)\}$

Not shown in the figure are possible transitions between states s_2, s_3 and s_4. This is to avoid cluttering the diagram. (There are no transitions from s_2, s_3 and s_4 to s_1 as once the deadline is reached that cannot be undone.) For simplicity we have ignored actions that are not relevant to the contract, and we have assumed that the performance of a (relevant) action in state s_1 leads to a state where the deadline has been reached. Obviously that could be adjusted, for instance by adding reflexive transitions $(s_1, \varepsilon_1, s_1)$ and $(s_1, \varepsilon_2, s_1)$ and introducing more states to allow for the work to be completed or an accident to have happened before the deadline is reached. The simple version of the example is sufficient for present purposes.

4.2. Introducing norms

So far there are no normative atoms in the picture. How can they be introduced? Ultimately we want to be able to answer the question: '*what should an agent do in a specific state of affairs?*'.

Norms codify the relation between the descriptive (non-normative) features of possible states and actions on the one hand and the normative evaluation of those states and actions on the other hand. In order to model norms adequately, we need a way of colouring a given (non-normative) LTS. This colouring moreover should be based on general descriptive terms rather than having to specify individual transitions and states explicitly one by one.

In this first example the norms that regulate states and transitions are context independent, that is, the normative status of every transition depends only on the type of action of which it is an instance. The situation in which that action is carried out and its results are not taken into account. That is not true in general.

For the example we want to define a colouring $\mathcal{M}_1$ of the states and transitions in the LTS $\mathcal{T}_1$. First it is natural to say that all actions that are unsafe are forbidden, i.e., that transitions of this type are never acceptable.

- $\|unsafe\_action\|^{\mathcal{M}_1} \subseteq \|red_t\|^{\mathcal{M}_1}$

That can be expressed by requiring the validity in $\mathcal{M}_1$ of the transition formula:

$$unsafe\_action \rightarrow red_t$$

or equivalently the validity in $\mathcal{M}_1$ of the state formula:

$$\underset{\rightarrow}{\Box}(unsafe\_action \rightarrow red_t)$$

Similarly we can say that any state in which a deadline has been reached without the work having been done, or any state in which an accident has happened, is unacceptable (red). That is validity in $\mathcal{M}_1$ of the state formulas:

$$deadline\_passed \wedge \neg work\_done \rightarrow red_s \tag{6}$$

$$accident \rightarrow red_s \tag{7}$$

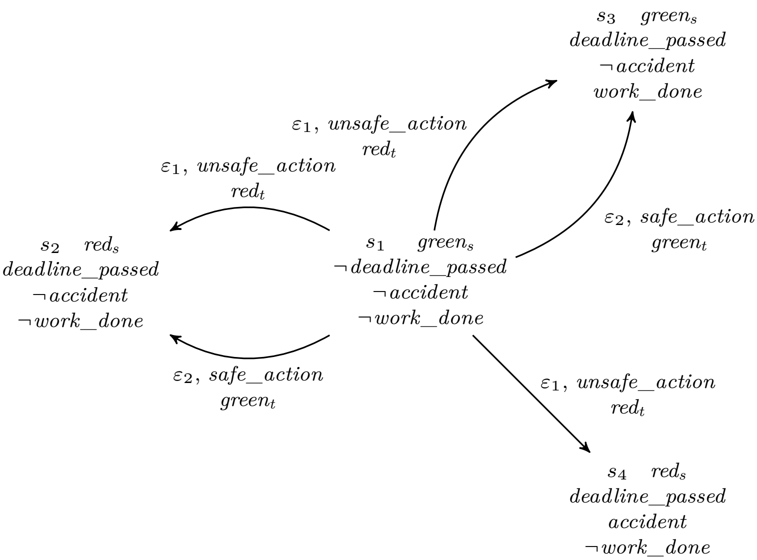

Figure 2. The coloured model $\mathcal{M}_1$ of Example 1

If being unsafe is the only reason for an action to be forbidden we can strengthen the inclusion to the following equalities:

- $\|red_t\|^{\mathcal{M}_1} = \|unsafe\_action\|^{\mathcal{M}_1}$
- $\|green_t\|^{\mathcal{M}_1} = \|safe\_action\|^{\mathcal{M}_1}$

For states similarly, if (6) and (7) are the only reasons for a state to be considered unacceptable (red), then s_2 and s_4 are red, whereas s_1 and s_3 are green:

- $\|red_s\|^{\mathcal{M}_1} =$
  $= \|(deadline\_passed \land \neg work\_done) \lor accident\_happened\|^{\mathcal{M}_1} =$
  $= \{s_2, s_4\}$
- $\|green_s\|^{\mathcal{M}_1} = \{s_1, s_3\}$

The model $\mathcal{M}_1$ with its normative layer is presented in Figure 2.

So here one can see, for instance:

- $\mathcal{M}, s_1 \models \underset{\rightarrow}{\Box}(green_t \rightarrow safe\_action)$, i.e., $\mathcal{M}, s_1 \models \underset{\rightarrow}{\mathsf{O}}\, safe\_action$

as well as, for example

- $\mathcal{M}, s_1 \models \neg accident \land \mathsf{O}^s \neg accident$
- $\mathcal{M}, s_4 \models accident \land \mathsf{O}^s \neg accident$
- $\mathcal{M}, s_2 \models deadline\_passed \land \neg work\_done \land \mathsf{O}^s(deadline\_passed \rightarrow work\_done)$

In the last example, the (global) obligation holds because

$$(\mathit{deadline\_passed} \wedge \neg \mathit{work\_done} \rightarrow \mathit{red}_s) \equiv (\mathit{green}_s \rightarrow (\mathit{deadline\_passed} \wedge \neg \mathit{work\_done})).$$

Of course there are many other equivalent ways of writing the same thing.

4.3. Generalisation

Given an (uncoloured) LTS $\mathcal{T}$ and a set Γ of transition and state formulas representing norms, we wish to construct a colouring $\mathcal{M}$ of $\mathcal{T}$ that satisfies Γ. We will call Γ the set of *explicit norms*.

The normative value ('colouring') of transitions can be specified by formulas of the following form:

$$\varphi \rightarrow \mathit{red}_t \tag{8}$$

expressing that if a transition satisfies φ then that transition is red. Intuitively φ *is meant here to be a rationale for the transition to be red.* The question arises what should be the structure of φ in these expressions. In principle we could allow φ to be any well formed transition formula, including formulas containing normative atoms red_t and green_t, such as, for example:

$$(0{:}\,\underset{\rightarrow}{\Box}(\psi \rightarrow \mathit{red}_t) \wedge \varphi) \rightarrow \mathit{red}_t$$

That is equivalent to

$$0{:}\,\underset{\rightarrow}{\Box}(\psi \rightarrow \mathit{red}_t) \rightarrow (\varphi \rightarrow \mathit{red}_t)$$

and so can be read as saying that a transition of type φ is red if all transitions of type ψ from the same initial state are red: roughly, φ is forbidden whenever ψ is forbidden.

However, in this paper we will restrict the form of φ in expression (8) in explicit norms to be transition formulas not containing the normative atoms red_t and green_t. (We allow the occurrence of operators 0: and 1: for reasons explained below. In that case φ must not contain the normative atoms red_s and green_s either.) We stress that these are restrictions on the form of expressions used as explicit norms, not on state and transition formulas that can be used to express properties of a coloured LTS.

In order to construct the colouring $\mathcal{M}$ of an LTS $\mathcal{T}$ we require that the transition formula (8) is valid in $\mathcal{M}$. That is equivalent to requiring that the state formula

$$\underset{\to}{\Box}(\varphi \to \mathit{red}_t)$$

is valid (true in all states) in $\mathcal{M}$. And that validity requirement is most conveniently expressed by means of the state formula:

$$\Box\, \underset{\to}{\Box}(\varphi \to \mathit{red}_t) \tag{9}$$

The operator $\Box$ reflects the idea that explicit norms are always general, i.e., hold in all states. It allows us to talk about (transition) norms being in effect at a particular state in a model, and from the technical point of view, it allows us to define a notion of local consequence between explicit norms and implied obligations and prohibitions. (See Section 6.)

In terms of the defined deontic operators $\underset{\to}{\mathsf{F}}\,\varphi =_{\mathrm{def}} \underset{\to}{\Box}(\varphi \to \mathit{red}_t)$ and $\underset{\to}{\mathsf{O}}\,\varphi =_{\mathrm{def}} \underset{\to}{\Box}(\mathit{green}_t \to \varphi)$ expression (9) may be written as

$$\Box\, \underset{\to}{\mathsf{F}}\,\varphi \quad \text{or} \quad \Box\, \underset{\to}{\mathsf{O}}\,\neg\varphi$$

In Section 5 we will generalise the form of these expressions to represent conditional norms.

Similarly, explicit state norms ('ought to be') for colouring states have the general form:

$$\Box\,(G \to \mathit{red}_s) \tag{10}$$

where G is a state formula not containing normative atoms red_s and green_s. G may contain the operator $\underset{\to}{\Box}$ (in which case it must not contain the atoms red_t and green_t).

In terms of the defined deontic operators F^s and O^s expression (10) can be written in the forms:

$$\mathsf{F}^s G \quad \text{or} \quad \mathsf{O}^s \neg G$$

Note that since the $\Box$ modality is of type S5, $\mathsf{F}^s G$ and $\mathsf{O}^s \neg G$ are already equivalent to $\Box\, \mathsf{F}^s G$ and $\Box\, \mathsf{O}^s \neg G$, respectively.

In the example, the explicit norms Γ_1 used to specify the colouring $\mathcal{M}_1$ of the LTS $\mathcal{T}_1$ can be written in the form:

- $\Box\, \underset{\to}{\mathsf{F}}\, \mathit{unsafe\_action}$
- $\mathsf{F}^s(\mathit{deadline\_passed} \land \neg \mathit{work\_done})$
- $\mathsf{F}^s\, \mathit{accident}$

The colouring $\mathcal{M}_1$ defined by these norms is the *most permissive* Γ_1-colouring of $\mathcal{T}_1$, that is (see Section 6): $\mathcal{M}_1 = \langle \mathcal{T}_1, S_{red}, R_{red} \rangle$ is the (unique) colouring of $\mathcal{T}$ in which all formulas of Γ_1 are valid and which minimises $S_{red} = h^n_S(red_s)$ and $R_{red} = h^n_R(red_t)$.

In this example we have

- $\mathcal{M}_1, s_1 \models \underset{\rightarrow}{\mathsf{O}}\, safe\_action$

Instead of naming the state s_1 explicitly we can also specify the state, or states, of interest by means of a general description[1]. Thus for example we have the following validity in $\mathcal{M}_1$:

- $\mathcal{M}_1 \models \neg\, work\_done \wedge \neg\, deadline\_passed \rightarrow \underset{\rightarrow}{\mathsf{O}}\, safe\_action$

We also have (since $\Box$ is the universal modality for states):

- $\mathcal{M}_1 \models \Box(\neg\, work\_done \wedge \neg\, deadline\_passed \rightarrow \underset{\rightarrow}{\mathsf{O}}\, safe\_action)$

As explained in Section 6 we can (if we wish) interpret that as a conditional norm implied by the explicit norms Γ_1 given the LTS $\mathcal{T}_1$.

4.4. Ought to do and ought to be: the *ggg* constraint

There is an intuition concerning the normative properties of actions and states that actions leading to bad (forbidden, illegal, undesirable, etc.) states are themselves bad. It is found in several works in deontic logic, most notably in Meyer's 'Dynamic deontic logic' [8]. There, an action is prohibited precisely when it leads to a prohibited state. It has been observed by several authors however, including [2, 14] as well as the present authors, that relationship is extremely problematic without qualification. It means that any action from a prohibited state which fails to restore that state to a non-prohibited state must itself be prohibited. So for example in a state in which there is an overdue debt (a prohibited state) the actions of paying one's taxes, or giving flowers to a neighbour, or even partially repaying the debt, are all prohibited because they all lead to a state in which a debt remains overdue (a prohibited state). That is simply not tenable.

A similar intuition underpins the *ggg* constraint referred to earlier but there it appears in a restricted form: the transfer from bad results to a bad transition only applies when an initial state is not already bad. There are two ways of looking at this. One way is as a kind of *well-formedness* principle: a green (permitted, acceptable, legal) transition

[1] This is how properties of a coloured LTS are investigated in the iCCalc implementation of the $n\mathcal{C}+$ language. The task of the system is to confirm validity or to find all counterexamples. See [10] for a detailed worked example.

in a green (permitted, acceptable, legal) state must always lead to a green (acceptable, legal, permitted) state (as also suggested by [2]). The other way to see it is as a restricted form of the Meyer constraint: a transition leading to a red state is red *but only* if the initial state is not also red.

In previous works the *ggg* constraint has been treated as a fixed feature of a coloured LTS. A coloured LTS which failed to satisfy the *ggg* constraint was not well defined. Here we treat it as an optional constraint, and we obtain it by defining a new transition atom $red_t^{\dagger}$. By transposition it follows from the *ggg* constraint that a transition starting from a green state and terminating in a red state must be red. Accordingly we define:

$$red_t^{\dagger} =_{\text{def}} \; red_t \vee (0{:}\,green_s \wedge 1{:}\,red_s) \qquad (11)$$

$$green_t^{\dagger} =_{\text{def}} \; \neg red_t^{\dagger}$$

or in full:

$$green_t^{\dagger} =_{\text{def}} \; green_t \wedge (0{:}\,red_s \vee 1{:}\,green_s) \qquad (12)$$

The definitions of the corresponding prohibition and obligation operators are as follows:

$$\underset{\rightarrow}{\underline{\mathsf{F}}}^{\dagger}\varphi \;=_{\text{def}}\; \underset{\rightarrow}{\Box}(\varphi \to red_t^{\dagger})$$

$$\underset{\rightarrow}{\underline{\mathsf{O}}}^{\dagger}\varphi \;=_{\text{def}}\; \underset{\rightarrow}{\Box}(green_t^{\dagger} \to \varphi)$$

$\underset{\rightarrow}{\underline{\mathsf{F}}}^{\dagger}$ is also cumulative and homogeneous. The following are valid:

$$\underset{\rightarrow}{\underline{\mathsf{F}}}^{\dagger}(\varphi \vee \psi) \leftrightarrow (\underset{\rightarrow}{\underline{\mathsf{F}}}^{\dagger}\varphi \wedge \underset{\rightarrow}{\underline{\mathsf{F}}}^{\dagger}\psi)$$

$$\underset{\rightarrow}{\underline{\mathsf{F}}}^{\dagger}(\varphi \wedge \neg\varphi)$$

$$\underset{\rightarrow}{\underline{\mathsf{F}}}^{\dagger}(\varphi \vee \neg\varphi) \leftrightarrow \underset{\rightarrow}{\Box}\, red_t^{\dagger}$$

In terms of $\underset{\rightarrow}{\underline{\mathsf{O}}}^{\dagger}$:

$$\underset{\rightarrow}{\underline{\mathsf{O}}}^{\dagger}(\varphi \wedge \psi) \leftrightarrow (\underset{\rightarrow}{\underline{\mathsf{O}}}^{\dagger}\varphi \wedge \underset{\rightarrow}{\underline{\mathsf{O}}}^{\dagger}\psi)$$

$$\underset{\rightarrow}{\underline{\mathsf{O}}}^{\dagger}(\varphi \vee \neg\varphi)$$

$$\underset{\rightarrow}{\underline{\mathsf{O}}}^{\dagger}(\varphi \wedge \neg\varphi) \leftrightarrow \underset{\rightarrow}{\Box}\, red_t^{\dagger}$$

And indeed, the properties of $\underset{\rightarrow}{\underline{\mathsf{O}}}^{\dagger}$ and $\underset{\rightarrow}{\underline{\mathsf{F}}}^{\dagger}$ must be the same as those of $\underset{\rightarrow}{\underline{\mathsf{O}}}$ and $\underset{\rightarrow}{\underline{\mathsf{F}}}$ since their definitions have exactly the same form and differ only in the propositional atoms in terms of which they are defined.

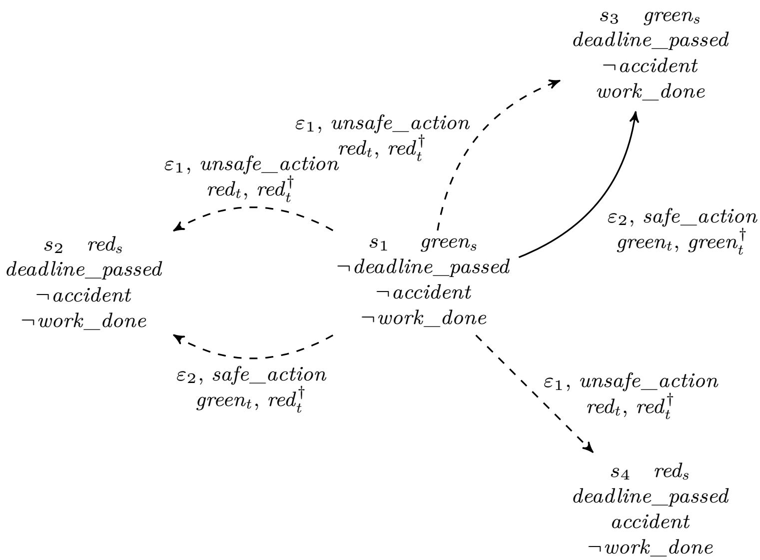

Figure 3. The solid (non-dashed) line shows the only $\mathit{green}_t^\dagger$ transition for Example 1

Obviously, from the definition, $\|\mathit{red}_t\|^{\mathcal{M}} \subseteq \|\mathit{red}_t^\dagger\|^{\mathcal{M}}$ (so $\mathit{red}_t \to \mathit{red}_t^\dagger$ is valid) and hence:

$$\models \underset{\rightarrow}{\mathsf{O}}\varphi \to \underset{\rightarrow}{\mathsf{O}}^\dagger\varphi$$

Figure 3 shows the $\mathit{red}_t^\dagger$ and $\mathit{green}_t^\dagger$ colouring for the example.

We have, for instance:

- $\mathcal{M}_1, s_1 \models \underset{\rightarrow}{\mathsf{F}}^\dagger 1{:}(\mathit{deadline\_passed} \wedge \neg\, \mathit{work\_done})$
- $\mathcal{M}_1, s_1 \models \underset{\rightarrow}{\mathsf{F}}^\dagger(\mathit{safe\_action} \wedge 1{:}(\mathit{deadline\_passed} \wedge \neg\, \mathit{work\_done}))$
- $\mathcal{M}_1, s_1 \models \underset{\rightarrow}{\mathsf{O}}^\dagger(\mathit{safe\_action} \wedge 1{:}(\mathit{deadline\_passed} \to \mathit{work\_done}))$

(The second is implied by the first. The third is equivalent to the second.)

In this particular example, the state s_1 of interest is green and so the *ggg* colouring is the same as would be obtained using Meyer's unrestricted version. That will not always be so, as later examples will show.

If we did want to formulate a Meyer-like constraint we could do so by defining another pair of normative transition atoms in similar style:

$$\mathit{red}_t^\star =_{\text{def}} \mathit{red}_t \vee 1{:}\mathit{red}_s$$
$$\mathit{green}_t^\star =_{\text{def}} \mathit{green}_t \wedge 1{:}\mathit{green}_s$$

Evidently $\mathit{green}_t^\star \equiv \neg\, \mathit{red}_t^\star$.

As a final illustration: it may be felt that the *ggg* constraint is too weak because it has nothing to say about transitions from a red initial state. One might take the view that if there is a choice between a transition leading to a red state and a transition from the same state leading to a green state, then the transition leading to the green state should be preferred.

Let us make that explicit. Let us strengthen (actually, complement) the *ggg* constraint, as follows. If there is a transition τ from a red state to a red state, and there exists an *alternative* transition τ' from that same initial red state to a green state, and that alternative τ' is green (not already red), then τ is red. Let us call it the $ggg^{\ddagger}$ constraint. It can be obtained by defining another pair of normative transition atoms:

$$red_t^{\ddagger} =_{\text{def}} \ red_t \vee (\, 0{:}red_s \wedge 1{:}red_s \wedge 0{:}\underset{\rightarrow}{\Diamond}(1{:}green_s \wedge green_t)\,)$$
$$green_t^{\ddagger} =_{\text{def}} \ \neg red_t^{\ddagger}$$

Note that this new $ggg^{\ddagger}$ constraint does not conflict with the *ggg* constraint at all. $ggg^{\ddagger}$ is applicable when the initial state is red. *ggg* is applicable when the initial state is green. So we can see the combination of $ggg^{\ddagger}$ and *ggg* as a corrected form of the Meyer-like constraint. The $ggg^{\ddagger}$ constraint deals with *recovery* from red states. The original *ggg* constraint deals only with transitions from green states.

There are some obvious relationships between these various forms of red and the corresponding deontic operators. For example, a transition cannot be both $red_t^{\dagger}$ and $red_t^{\ddagger}$, unless it is already red_t. In other words:

$$\models red_t^{\dagger} \wedge red_t^{\ddagger} \rightarrow red_t$$

It is our view that the *ggg* constraint expresses a fundamental well-formedness principle that is not readily challenged. Although we have left it optional, in that $red_t^{\dagger}$ is defined separately from red_t, we expect that it will be $red_t^{\dagger}$ and $green_t^{\dagger}$ transitions that will be of primary interest. In contrast, the complementary $ggg^{\ddagger}$ constraint can be challenged and might not be appropriate in all circumstances, particularly in the multi-agent context (which we are not discussing in this paper).

One might prefer therefore to leave the $ggg^{\ddagger}$ constraint optional, and to define another separate form of obligation that combines the effects of *ggg* and $ggg^{\ddagger}$, as

$$\underset{\rightarrow}{\Box}((green_t^{\dagger} \wedge green_t^{\ddagger}) \rightarrow \varphi)$$

We have not given a symbol to this form of obligation in this paper in order to avoid a proliferation of notations and we do not record its properties. They are very readily reconstructed: again, we have a form of obligation whose definition has exactly the same form as others and which differs only in the propositional atom used to define it.

4.5. Complete obligations

We now want to introduce a form of obligation that determines at each state the complete set of all acceptable transitions from that state. This form of obligation will not be used in the formulation of explicit norms but rather provides a means of guiding actions given a set of explicit norms already specified. This form of obligation is a local one: we are looking for a set of transitions from which an agent can choose at a given state.

We define:

$$\mathsf{Obl}\,\varphi \;=_{\text{def}}\; \underset{\rightarrow}{\Box}(\varphi \leftrightarrow \mathit{green}_t)$$

where φ is a non-normative transition formula, i.e., a transition formula without normative atoms red_t and green_t (and red_s and green_s).

The corresponding satisfaction conditions are:

$$\mathcal{M}, s \models \mathsf{Obl}\,\varphi \quad \text{iff} \quad \mathit{exe}(s) \cap \|\varphi\|^{\mathcal{M}} = \mathit{exe}(s) \cap \|\mathit{green}_t\|^{\mathcal{M}}$$

$\mathit{exe}(s)$ is the set of possible transitions from state s. We can say then that the content of $\mathit{exe}(s) \cap \|\mathit{green}_s\|^{\mathcal{M}}$ is the complement in $\mathit{exe}(s)$ of the most general (in the sense of their extensions) transition type prohibited in s. So then φ in $\mathsf{Obl}\,\varphi$ denotes the *most specific* transition type that is obligatory in s.

In the example we have:

- $\mathcal{M}_1, s_1 \models \mathsf{Obl}\,\mathit{safe\_action}$

One can see immediately from the definitions that the following two properties are valid:

$$\mathsf{Obl}\,\varphi \rightarrow \underset{\rightarrow}{\mathsf{O}}\,\varphi$$

$$(\mathsf{Obl}\,\varphi \wedge \underset{\rightarrow}{\mathsf{O}}\,\psi) \rightarrow \underset{\rightarrow}{\Box}(\varphi \rightarrow \psi)$$

The first is just $\models \underset{\rightarrow}{\Box}(\mathit{green}_t \leftrightarrow \varphi) \rightarrow \underset{\rightarrow}{\Box}(\mathit{green}_t \rightarrow \varphi)$. The second is because $\models (\mathsf{Obl}\,\varphi \wedge \underset{\rightarrow}{\mathsf{O}}\,\psi) \rightarrow (\underset{\rightarrow}{\Box}(\varphi \rightarrow \mathit{green}_t) \wedge \underset{\rightarrow}{\Box}(\mathit{green}_t \rightarrow \psi))$. The two together express that $\mathsf{Obl}\,\varphi$ represents the most specific obligation φ.

Clearly Obl picks out a unique set of transitions at any state:

$$\models \ (\mathsf{Obl}\,\varphi \wedge \mathsf{Obl}\,\psi) \rightarrow (\varphi \leftrightarrow \psi)$$

And unlike $\underrightarrow{\mathsf{O}}$:

$$\not\models \ \underrightarrow{\Box}\varphi \rightarrow \mathsf{Obl}\,\varphi$$

The idea is that when $\mathsf{Obl}\,\varphi$ holds and φ is an expression in disjunctive normal form — $\varphi = \varphi_1 \vee \ldots \vee \varphi_n$ where each disjunct is a (consistent) conjunction of literals of the form α or $\neg\alpha$ where α is an atom of P_R or of the form $0{:}p$ or $0{:}\neg p$ or $1{:}p$ or $1{:}\neg p$ where p is an atom of P_S — then every disjunct φ_i of φ is a description of a package of actions that must/must not be performed in the current state. Atoms of P_R not appearing in φ_i are optional. Performance of any disjunct satisfies the obligation. If there is more than one disjunct in φ then the obligation is to perform one of them.

Suppose for example that there is a single explicit norm $\underrightarrow{\mathsf{O}}\neg(\mathit{drink} \wedge \mathit{drive})$. Then the following holds:

$$\mathsf{Obl}\,(\neg\,\mathit{drink} \vee \neg\,\mathit{drive}) \tag{13}$$

The disjunct $\neg\,\mathit{drink}$ represents the package of actions not drinking, and (optionally) either driving or not driving; $\neg\,\mathit{drive}$ represents the package of actions not driving, and (optionally) either drinking or not drinking. Expression (13) can also be written equivalently in the form

$$\mathsf{Obl}\,((\mathit{drink} \wedge \neg\,\mathit{drive}) \vee (\mathit{drive} \wedge \neg\,\mathit{drink}) \vee (\neg\,\mathit{drink} \wedge \neg\,\mathit{drive}))$$

In this form the disjuncts are mutually exclusive. The obligation is to perform exactly one of them.

Obl can be combined with the *ggg* constraint in order to obtain the corresponding form of most specific 'complete obligation'[2]:

$$\mathsf{Obl}^{\dagger}\varphi \ =_{\mathrm{def}} \ \underrightarrow{\Box}(\varphi \leftrightarrow \mathit{green}_t^{\dagger})$$

It has the following satisfaction condition (cf. the definition (12) of $\mathit{green}_t^{\dagger}$):

$$\begin{aligned}\mathcal{M}, s \models \mathsf{Obl}^{\dagger}\varphi \ \text{ iff } \ & \mathit{exe}(s) \cap \|\varphi\|^{\mathcal{M}} = \mathit{exe}(s) \cap \|\mathit{green}_t\|^{\mathcal{M}} \cap \\ & \{\tau \in R \mid \mathit{prev}(\tau) \in \|\mathit{red}_s\|^{\mathcal{M}} \text{ or } \mathit{post}(\tau) \in \|\mathit{green}_s\|^{\mathcal{M}}\}\end{aligned}$$

The relationships between $\mathsf{Obl}^{\dagger}$ and $\underrightarrow{\mathsf{O}}^{\dagger}$ are just as for Obl and $\underrightarrow{\mathsf{O}}$: the definitions have exactly the same form but with $\mathit{green}_t^{\dagger}$ for green_t.

[2] The construction is similar to the account in [6, 13].

In our example we have:

- $\mathcal{M}_1, s_1 \models \mathsf{Obl}^\dagger(\mathit{safe\_action} \wedge 1{:}(\mathit{deadline\_passed} \wedge \mathit{work\_done} \wedge \neg\, \mathit{accident}))$

Since $\mathit{safe\_action} \to 1{:}(\mathit{deadline\_passed} \wedge \neg\, \mathit{accident})$ is valid in the LTS $\mathcal{T}_1$, and therefore also in $\mathcal{M}_1$, we also have:

- $\mathcal{M}_1, s_1 \models \mathsf{Obl}^\dagger(\mathit{safe\_action} \wedge 1{:}\mathit{work\_done})$

Again, instead of having to name a state of interest (here s_1) explicitly we can also specify it descriptively. In the model $\mathcal{M}_1$:

- $\mathcal{M}_1 \models \Box(\neg\, \mathit{work\_done} \wedge \neg\, \mathit{deadline\_passed} \to \mathsf{Obl}^\dagger(\mathit{safe\_action} \wedge 1{:}\mathit{work\_done}))$

Of course we also have all obligations implied by the most specific 'complete' one, and therefore among other things:

- $\mathcal{M}_1 \models \Box(\neg\, \mathit{work\_done} \wedge \neg\, \mathit{deadline\_passed} \to \underrightarrow{\mathsf{O}}^\dagger(\mathit{safe\_action} \wedge 1{:}\mathit{work\_done}))$
- $\mathcal{M}_1 \models \Box(\neg\, \mathit{work\_done} \wedge \neg\, \mathit{deadline\_passed} \to \underrightarrow{\mathsf{O}}^\dagger\, \mathit{safe\_action})$
- $\mathcal{M}_1 \models \Box(\neg\, \mathit{work\_done} \wedge \neg\, \mathit{deadline\_passed} \to \underrightarrow{\mathsf{O}}^\dagger\, 1{:}\mathit{work\_done})$

and indeed, trivially,

$$\mathcal{M}_1 \models \Box(\neg\, \mathit{work\_done} \wedge \neg\, \mathit{deadline\_passed} \to \underrightarrow{\mathsf{O}}^\dagger \top).$$

4.6. Relationships between the deontic operators

Some of the following properties were noted earlier. They are repeated here for ease of reference. The properties of the individual deontic operators are the same for each variant since their definitions have the same form and differ only in the normative propositional atom in terms of which they are defined.

$$\underrightarrow{\mathsf{O}}\,\varphi \to \underrightarrow{\mathsf{O}}^\dagger\varphi \tag{14}$$

$$\mathsf{O}^s F \wedge \mathit{green}_s \to \underrightarrow{\mathsf{O}}^\dagger 1{:}F \tag{15}$$

$$\mathsf{Obl}\,\varphi \to \underrightarrow{\mathsf{O}}\,\varphi \tag{16}$$

$$(\mathsf{Obl}\,\varphi \wedge \underrightarrow{\mathsf{O}}\,\psi) \to \underrightarrow{\Box}(\varphi \to \psi) \tag{17}$$

$$\mathsf{Obl}^\dagger\varphi \to \underrightarrow{\mathsf{O}}^\dagger\varphi \tag{18}$$

$$(\mathsf{Obl}^\dagger\varphi \wedge \underrightarrow{\mathsf{O}}^\dagger\psi) \to \underrightarrow{\Box}(\varphi \to \psi) \tag{19}$$

$$(\mathsf{Obl}^\dagger\varphi \wedge \underrightarrow{\mathsf{O}}\,\psi) \to \underrightarrow{\Box}(\varphi \to \psi) \tag{20}$$

$$(\mathsf{Obl}^\dagger\varphi \wedge \mathsf{Obl}\,\psi) \to \underrightarrow{\Box}(\varphi \to \psi) \tag{21}$$

Note that $\not\models \mathsf{Obl}^\dagger \varphi \to \underset{\rightarrow}{\mathsf{O}}\, \varphi$. It is easy to construct a counterexample: a $green_s$ state with a single $green_t$ transition leading to a red_s state can have $\underset{\rightarrow}{\Box}(\varphi \leftrightarrow green_t^\dagger) \wedge \underset{\rightarrow}{\Diamond}(green_t \wedge \neg\varphi)$ true.

Properties (14) and (15) are from the *ggg* constraint. (14) was noted earlier and is valid because $\models green_t^\dagger \to green_t$. Property (15) is shown below. Formulas (16) and (17) say that Obl is the most specific $\underset{\rightarrow}{\mathsf{O}}$, and (18) and (19) that $\mathsf{Obl}^\dagger$ is the most specific $\underset{\rightarrow}{\mathsf{O}}^\dagger$. Property (20) is from $\models green_t^\dagger \to green_t$. (21) follows from (20) and (16).

To see that property (15) is valid: note first that $(green_s \to \underset{\rightarrow}{\mathsf{O}}^\dagger 1{:}F) \equiv \underset{\rightarrow}{\mathsf{O}}^\dagger(0{:}green_s \to 1{:}F)$ follows from property (3) (see also the discussion of conditional norms in Section 5 below). So it is enough to show the validity of

$$\mathsf{O}^s F \to \underset{\rightarrow}{\mathsf{O}}^\dagger(0{:}green_s \to 1{:}F)$$

that is, $\Box(green_s \to F) \to \underset{\rightarrow}{\Box}(green_t^\dagger \to (0{:}green_s \to 1{:}F)$, which is equivalent to

$$\Box(green_s \to F) \to \underset{\rightarrow}{\Box}(green_t^\dagger \wedge 0{:}green_s \to 1{:}F)$$

Now, from the definition:

$$(green_t^\dagger \wedge 0{:}green_s) \equiv (green_t^\dagger \wedge 0{:}green_s \wedge 1{:}green_s)$$

and hence

$$\models \underset{\rightarrow}{\Box}(green_t^\dagger \wedge 0{:}green_s \to 1{:}green_s)$$

It just remains to show:

$$\Box(green_s \to F) \to \underset{\rightarrow}{\Box}(1{:}green_s \to 1{:}F)$$

That follows from the general property (2) which was noted earlier.

5. Conditional norms

5.1. Example

In Section 4 we discussed how unconditional norms can be represented in the coloured LTS framework. In this section we focus on situations where the colouring of transitions is context dependent and unconditional norms are not adequate.

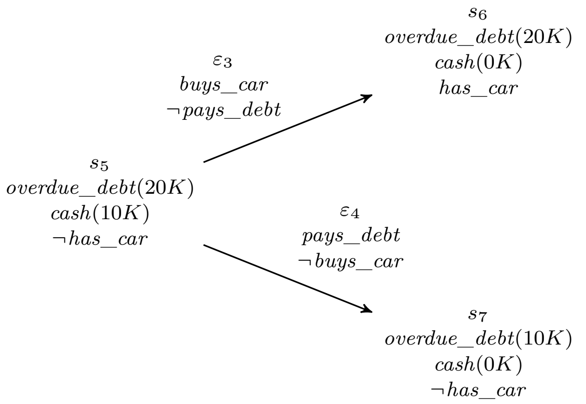

Figure 4. The initial model $\mathcal{T}_2$ of Example 2.

Example 2. Consider the situation (state s_5) of a person having an overdue debt of £20,000 and £10,000 to spend. She can buy a car (action ε_3) or pay off half of her debt (action ε_4). As a consequence, in state s_6 she has a car and an overdue debt of £20,000 and in state s_7 she has no car but an overdue debt of £10,000. (See Figure 4.) ⊣

In this example there is a social (and also to some extent legal) rule that debts should be paid off on time. Thus, any state in which there are any (significant) overdue debts is red. So we have the following pair of explicit norms (two, because in our simplified representation of the example there are two possible values $10K$ and $20K$ of an overdue debt):

$$\begin{aligned} &\mathsf{F}^s\, overdue\_debt(20K) \\ &\mathsf{F}^s\, overdue\_debt(10K) \end{aligned} \tag{22}$$

With just these explicit norms the most permissive colouring of the LTS has:

- $\|overdue\_debt(20K) \lor overdue\_debt(10K)\| = \|red_s\| = \{s_5, s_6, s_7\}$
- $\|red_t\| = \{\}$
- $\|green_t\| = \{(s_5, \varepsilon_4, s_7), (s_5, \varepsilon_3, s_6)\} = \|pays\_debt \lor buys\_car\|$.

All three states (in the small fragment of the example we are considering) are coloured red because globally it is unacceptable (forbidden) to have an overdue debt. Both transitions are coloured green because so far we have not said otherwise: in principle there is nothing wrong in buying a car. However, we might reasonably take the view that in s_5, the agent, although unable to pay off the debt completely, should nevertheless pay off what part of it she can instead of spending the available money on (unnecessary) goods. The *ggg* constraint does not colour such transitions $red_t^\dagger$ since all states are red_s.

5.2. Conditional transition norms

It is not *action types* such as buying a car or paying a debt that are coloured red/green in this framework but *transitions*, which are action types in context. Conditional norms allow transitions to be identified by reference to the features of their initial (and in principle final) states.

The general form of a conditional explicit norm is:

$$\Box(F \to \underset{\rightarrow}{\Box}(\varphi \to red_t)) \tag{23}$$

where F is a state formula and φ is a transition formula, neither of which contain the normative atoms red_s, $green_s$ and red_t, $green_t$.

In terms of the defined deontic operators $\underset{\rightarrow}{\mathsf{F}}$ and $\underset{\rightarrow}{\mathsf{O}}$ expression (23) may be written in the forms

$$\Box(F \to \underset{\rightarrow}{\mathsf{F}}\varphi)$$
$$\Box(F \to \underset{\rightarrow}{\mathsf{O}}\neg\varphi)$$

Expressions $\Box\underset{\rightarrow}{\mathsf{F}}\varphi$ and $\Box\underset{\rightarrow}{\mathsf{O}}\neg\varphi$ used earlier for unconditional explicit norms are equivalent to

$$\Box(\top \to \underset{\rightarrow}{\mathsf{F}}\varphi)$$
$$\Box(\top \to \underset{\rightarrow}{\mathsf{O}}\neg\varphi)$$

indicating that, as expected, unconditional norms are a special case.

In our example we can formulate the following norm: "in a situation where an agent has an overdue debt and money to spend, she should not avoid paying off her debt (as much as she is able)". Thus, for X ranging over $10K$ and $20K$:

$$\Box(overdue\_debt(X) \land \neg cash(0K) \to \underset{\rightarrow}{\mathsf{F}}\neg pays\_debt) \tag{24}$$

The norm (with $X = 20K$) is in force in state s_5 since its preconditions are true there.

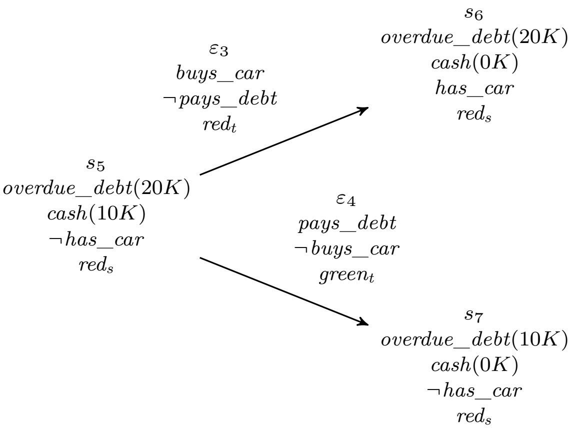

Figure 5. The coloured model $\mathcal{M}_2$ of Example 2

Let $\mathcal{M}_2$ be the most permissive colouring of $\mathcal{T}_2$ with explicit norms (22) and (24). Then:

$$\mathcal{M}_2, s_5 \models \underset{\rightarrow}{\mathsf{F}} \neg \mathit{pays\_debt}, \quad \text{i.e.} \quad \mathcal{M}_2, s_5 \models \underset{\rightarrow}{\Box} (\neg \mathit{pays\_debt} \rightarrow \mathit{red}_t)$$

When $\mathcal{M}_2$ is the most permissive colouring:

$$\|\mathit{red}_t\|^{\mathcal{M}_2} = \{(s_5, \varepsilon_3, s_6)\}$$

The coloured model $\mathcal{M}_2$ is shown in Figure 5.

Notice that in $\mathcal{M}_2$ at the state s_5 we have:

$$\mathcal{M}_2, s_5 \models \underset{\rightarrow}{\Box} (\mathit{buys\_car} \rightarrow \mathit{red}_t)$$

And indeed we have a (pair of) implied norms. The following are valid in $\mathcal{M}_2$ for $X = 10K$ and $X = 20K$:

$$\Box (\mathit{overdue\_debt}(X) \land \neg \mathit{cash}(0K) \rightarrow \underset{\rightarrow}{\mathsf{F}} \mathit{buys\_car})$$

The *ggg* constraint has no effect because every state in this simple version of the example is red. The example illustrates however why the stronger version, where a transition is coloured red if it leads to a red state, is unacceptable in general. In this example, a transition in which the agent pays off part of her debt instead of buying a car would be forbidden, coloured red because in the resulting state there is still an outstanding overdue debt. The *ggg* constraint in contrast does not apply when the initial state is red, precisely to avoid this kind of conclusion.

Finally, recall that the following is a general property (3) of every (coloured and uncoloured) LTS:

$$(F \to \underset{\rightarrow}{\Box}\varphi) \;\equiv\; \underset{\rightarrow}{\Box}(0{:}F \to \varphi)$$

The general form (23) of a conditional norm can thus be written equivalently as:

$$\Box\,\underset{\rightarrow}{\Box}(0{:}F \wedge \varphi \to \mathit{red}_t)$$

which is a special case of an unconditional norm (9). That again emphasises that what is coloured red/green are not action types but transitions.

In terms of the defined obligation operator $\underset{\rightarrow}{\mathsf{O}}$ we have:

$$\Box\,(F \to \underset{\rightarrow}{\mathsf{O}}\,\varphi) \equiv \Box\,\underset{\rightarrow}{\mathsf{O}}\,(0{:}F \to \varphi)$$

and more generally

$$\Box\,(F \wedge G \to \underset{\rightarrow}{\mathsf{O}}\,\varphi) \equiv \Box\,(F \to \underset{\rightarrow}{\mathsf{O}}\,(0{:}G \to \varphi))$$

Although these alternative forms are perhaps less natural than (23), for the sake of flexibility we see no reason to disallow them as a form of explicit norm. The expression φ moreover are allowed to contain the 1: operator. For example, suppose that in Example 1, we wanted to say (hypothetically, for the sake of an example) that an *unsafe_action* is forbidden (unacceptable, red) if it leads to an accident but not necessarily otherwise. That would be expressed by the explicit norm:

$$\Box\,\underset{\rightarrow}{\Box}(\mathit{unsafe\_action} \wedge 1{:}\mathit{accident} \to \mathit{red}_t), \quad \text{i.e.}$$

$$\Box\,\underset{\rightarrow}{\mathsf{F}}\,(\mathit{unsafe\_action} \wedge 1{:}\mathit{accident})$$

Whether it is *reasonable* to formulate norms such as this, where what is forbidden depends not only on what is true in the current state but also on the eventual future outcome, is another matter. Questions such as these are discussed elsewhere [3, 11] under the heading of 'absence of moral luck'. That is another kind of *rationality principle* that we might require of a well-formed and effective set of norms.

To take one last example, the transition norm

$$\underset{\rightarrow}{\mathsf{O}}\,\neg 1{:}(\mathit{window\_open} \wedge \mathit{heat\_on})$$

requires that, whatever action is taken in the current state, it must not result in a state where the window is open (*window_open*) and the heating is on (*heat_on*). The state norm

$$\mathsf{O}^s \neg(\mathit{window\_open} \wedge \mathit{heat\_on})$$

is slightly weaker. With the *ggg* constraint, it has the same effect as the transition norm when the current state is green, but it does not prescribe what should be done in a red state where the window is already open and the heating is on.

5.3. Conditional state obligations

In this framework, state norms/colourings are *global* properties of a model. If F is obligatory at a state s in a model $\mathcal{M}$ then F is obligatory at every state in $\mathcal{M}$. An expression such as 'if F then it ought to be G' is not expressed by the formula

$$\Box(F \to \mathsf{O}^s G)$$

equivalently $\Box(F \to \Box(\neg G \to \mathit{red}_s))$, but by the formula

$$\mathsf{O}^s(F \to G)$$

equivalently $\Box((F \wedge \neg G) \to \mathit{red}_s)$ or $\Box(F \to (\neg G \to \mathit{red}_s))$. $\Box(F \to \mathsf{O}^s G)$ would say something different: that if F is true at any state in a model $\mathcal{M}$ then $\mathsf{O}^s G$ is true at every state in $\mathcal{M}$. That is not a form of conditional obligation as the term is ordinarily understood.

As defined in this framework, the state obligation operator O^s does not satisfy what is often called 'factual detachment' in deontic logic:

$$\not\models\; F \wedge \mathsf{O}^s(F \to G) \to \mathsf{O}^s G$$

It has the property called 'deontic detachment':

$$\models\; \mathsf{O}^s F \wedge \mathsf{O}^s(F \to G) \to \mathsf{O}^s G$$

(It also satisfies what is sometimes called 'strong factual detachment':

$$\models\; \Box F \wedge \mathsf{O}^s(F \to G) \to \mathsf{O}^s G$$

but when $\Box$ is merely the universal modality for states, as here, this is of little significance.)

For comparison, the transition obligation $\underset{\rightarrow}{\mathsf{O}}$ satisfies deontic detachment:

$$\models \ \underset{\rightarrow}{\mathsf{O}}\varphi \wedge \underset{\rightarrow}{\mathsf{O}}(\varphi \rightarrow \psi) \rightarrow \underset{\rightarrow}{\mathsf{O}}\psi$$

but there is also a form of factual detachment, in that:

$$\models \ F \wedge \Box(F \rightarrow \underset{\rightarrow}{\mathsf{O}}\varphi) \rightarrow \underset{\rightarrow}{\mathsf{O}}\varphi$$

6. Norms and most permissive models

Let us summarize. We are given an LTS

$$\mathcal{T} = \langle S, A, R, \mathit{prev}, \mathit{post}, \mathit{label}, h_S, h_R \rangle.$$

How $\mathcal{T}$ is defined and constructed does not matter for the purposes of this paper. We assume it is given. $\mathcal{T}$ is not coloured. It provides an interpretation of all non-normative propositional atoms. Let us call this language $\mathcal{L}$. Let $\mathcal{L}^n$ be the language $\mathcal{L}$ extended with propositional atoms red_s and red_t (and green_s and green_t), any defined atoms such as $\mathit{red}_t^\dagger$ and $\mathit{green}_t^\dagger$, and the defined operators O^s and F^s, $\underset{\rightarrow}{\mathsf{O}}$ and $\underset{\rightarrow}{\mathsf{F}}$, $\underset{\rightarrow}{\mathsf{O}}^\dagger$, Obl and $\mathsf{Obl}^\dagger$, and so on.

Now we have a set Γ of expression of $\mathcal{L}^n$ representing *explicit norms*. Essentially these Γ are conditional expressions (rules) defining red_s and red_t. They will be used to construct a coloured model $\mathcal{M} = \langle \mathcal{T}, h_S^n, h_R^n \rangle$ of the LTS $\mathcal{T}$ on which we can evaluate formulas of $\mathcal{L}^n$. With certain (natural) restrictions on the form of Γ we can ensure that this colouring $\mathcal{M}$ of $\mathcal{T}$ is well-defined and unique.

Explicit norms Γ are state formulas of $\mathcal{L}$. They are of two kinds:

State norms:

$$\Box(F \rightarrow \mathit{red}_s), \qquad \mathsf{O}^s F, \qquad \mathsf{F}^s F \tag{25}$$

Transition norms:

$$\Box(F \rightarrow \underset{\rightarrow}{\Box}(\varphi \rightarrow \mathit{red}_t)), \qquad \Box(F \rightarrow \underset{\rightarrow}{\mathsf{O}}\varphi), \qquad \Box(F \rightarrow \underset{\rightarrow}{\mathsf{F}}\varphi) \tag{26}$$

F and φ are state and transition formulas of $\mathcal{L}$, respectively (and hence do not contain any of the normative propositional atoms in $\mathcal{L}^n$). In a transition norm, if $F \equiv \top$ it may be omitted. (The resulting expression is then an unconditional norm.)

F and φ are state and transition formulas of $\mathcal{L}$, respectively (and hence do not contain any of the normative propositional atoms in $\mathcal{L}^n$). In a transition norm, if $F \equiv \top$ it may be omitted. (The resulting expression is then an unconditional norm.)

We expect that usually F and φ will be truth-functional formulas of propositional atoms only. However, for flexibility, we allow the operators $\Box$ and $\underset{\rightarrow}{\Box}$ and 0: and 1: to appear in F and φ.

A colouring $\mathcal{M} = \langle \mathcal{T}, h^n_S, h^n_R \rangle$ of $\mathcal{T}$ satisfies the explicit norms Γ when every expression in Γ is valid in $\mathcal{M}$. We write $\mathcal{M} \models \Gamma$.

Since satisfaction of a set of explicit norms is defined in terms of validity, we could have allowed explicit norms to be written without the universal state modality $\Box$. We have retained it to emphasise that explicit norms are global properties of a model.

Most permissive models. A model $\mathcal{M} = \langle \mathcal{T}, h^n_S, h^n_R \rangle$ is *more permissive than* a model $\mathcal{M}' = \langle \mathcal{T}, h^n_S{}', h^n_R{}' \rangle$ when $h^n_S(\mathit{red}_s) \subseteq h^n_S{}'(\mathit{red}_s)$ and $h^n_R(\mathit{red}_t) \subseteq h^n_R{}'(\mathit{red}_t)$.

$\mathcal{M} = \langle \mathcal{T}, h^n_S, h^n_R \rangle$ is the *most permissive* Γ-colouring of $\mathcal{T}$ when (i) $\mathcal{M}$ satisfies Γ and (ii) for any colouring $\mathcal{M}' = \langle \mathcal{T}, h^n_S{}', h^n_R{}' \rangle$ of $\mathcal{T}$ which satisfies Γ, $h^n_S(\mathit{red}_s) \subseteq h^n_S{}'(\mathit{red}_s)$ and $h^n_R(\mathit{red}_t) \subseteq h^n_R{}'(\mathit{red}_t)$. We write $\mathcal{M}_{\mathcal{T},\Gamma}$ for the most permissive Γ colouring of $\mathcal{T}$. It is the colouring of $\mathcal{T}$ that satisfies Γ and (uniquely) minimises $S_{\mathit{red}} = h^n_S(\mathit{red}_s)$ and $R_{\mathit{red}} = h^n_R(\mathit{red}_t)$.

$\mathcal{M}_{\mathcal{T},\Gamma}$ is well-defined and unique. One can see this by encoding the LTS $\mathcal{T}$ together with the explicit norms Γ as a logic program: $\mathcal{T}$ can be encoded as a (not necessarily finite) set of atoms, one atom for each state and one atom for each transition in $\mathcal{T}$, and the explicit norms Γ as rules defining red_s and red_t in terms of these atoms. That logic program is definite and so has a least (unique minimal) model, which encodes $\mathcal{M}_{\mathcal{T},\Gamma}$. (These are properties of a definite logic program found in any standard text on logic programming.) Moreover, if colourings $\langle \mathcal{T}, h^n_S, h^n_R \rangle$ and $\langle \mathcal{T}, h^n_S{}', h^n_R{}' \rangle$ satisfy Γ then the colouring $\langle \mathcal{T}, h^n_S(\mathit{red}_s) \cap h^n_S{}'(\mathit{red}_s), h^n_R(\mathit{red}_t) \cap h^n_R{}'(\mathit{red}_t) \rangle$ satisfies Γ.

Since the definition (11) of the *ggg* atom $\mathit{red}^\dagger_t$ can also be encoded as definite rules, the most permissive colouring $\mathcal{M}_{\mathcal{T},\Gamma} = \langle \mathcal{T}, h^n_S, h^n_R \rangle$ also minimises $h^n_R(\mathit{red}^\dagger_t)$.

Many variations are possible. In particular, one could dispense with the distinction between red_t and $\mathit{red}^\dagger_t$, as in $n\mathcal{C}+$ [3, 12]. The definition

(11) then takes the form of a constraint (the *ggg* constraint), and is satisfied by $\mathcal{M}_{\mathcal{T},\Gamma}$.

We stress again that the restrictions imposed on the form of explicit norms only apply when those expressions are used as explicit norms. There are no restrictions on formulas of $\mathcal{L}^n$ to be evaluated on $\mathcal{M}_{\mathcal{T},\Gamma}$. Properties of $\mathcal{M}_{\mathcal{T},\Gamma}$ can be investigated without having to name states and transitions explicitly. We can describe states and transitions of interest by means of a state formula F or transition formula φ expressing the relevant properties, and then determine whether $\mathcal{M}_{\mathcal{T},\Gamma} \models F \to G$ or $\mathcal{M}_{\mathcal{T},\Gamma} \models 0{:}F \to \psi$ or $\mathcal{M}_{\mathcal{T},\Gamma} \models \varphi \to \psi$ where formulas G or ψ express whatever it is we wish to know. When

$$\mathcal{M}_{\mathcal{T},\Gamma} \models \Box(F \to \underset{\to}{\mathsf{O}}\varphi)$$

we can say that $\Box(F \to \underset{\to}{\mathsf{O}}\varphi)$ is a norm implied by Γ on LTS $\mathcal{T}$, and similarly for the other forms (25) and (26).

For the multi-agent case, the language $\mathcal{L}^n$ is readily extended to deal with agent-specific norms as in [3, 10]. The atoms red_s and red_t then represent impersonal (or 'system') norms, and atoms $\mathit{red}_s(x)$, $\mathit{red}_t(x)$ and $\mathit{red}_t^{\dagger}(x)$ for each agent x what is unacceptable (red) from the point of view of agent x, with the corresponding defined operators relativised to agents. The methods for each agent-specific colouring work in exactly the same way as those presented in this paper. The language $\mathcal{L}^n$ can also be extended by agency ('sees to it that', 'brings it about that') operators [10, 11]. That allows one to express, for example, whether a specific agent x was responsible for a given transition being red_t or for the recovery from a red_s state to a green_s state. Examples are provided in the works cited above. The colouring methods are not affected by that extension.

Extensions. The framework itself can be extended in several ways. First, we can generalise the form of explicit norms. One easy way is to allow permissions (rules defining green_s and green_t) to act as exceptions to general rules defining red_s and red_t. $n\mathcal{C}+$ has some support for mechanisms of this kind. With reasonable restrictions, we can still guarantee that the most permissive colouring is unique. More generally, we are currently looking at allowing defeasible explicit norms together with a priority ordering between them to resolve conflicts. In that case the most permissive colourings are not necessarily unique, just as in general there are multiple answer sets for a logic program or multiple extensions in default logic. More generally still, we might look at moving away from most permissive colourings, where states and transitions are always green by default unless coloured red by the explicit norms, to a more flexible framework which treats red and green more symmetrically. (The last question is perhaps of more theoretical than practical value.)

In general then: given an uncoloured LTS $\mathcal{T}$ we can define a deducibility relation $\vdash_{\mathcal{T}}$ as follows. A formula φ of $\mathcal{L}^n$ is valid in $\mathcal{T}$, $\models_{\mathcal{T}} \varphi$, if φ is valid in all models that are possible colourings of $\mathcal{T}$ of some specified kind (for instance, but not necessarily, the most permissive ones). In other words, the uncoloured LTS $\mathcal{T}$ can be seen as defining a class of coloured models, namely, all models that are possible colourings (of a specified kind) of $\mathcal{T}$. Deducibility is then defined as usual: $\Gamma \vdash_{\mathcal{T}} \varphi$ if $\models_{\mathcal{T}} (\varphi_1 \wedge \ldots \wedge \varphi_n) \to \varphi$ for some $\{\varphi_1, \ldots, \varphi_n\} \subseteq \Gamma$. The most permissive colourings for the restricted forms of explicit norms Γ presented in this paper are a very special case. This second account is much more general however since in principle it can work on any set of expressions Γ of $\mathcal{L}^n$, not only those which represent norms.

7. Conclusions

There are two main contributions in this paper. Unlike in $n\mathcal{C}+$, where $n\mathcal{C}+$ rules are written to specify which states and transitions are green/red as part of defining a coloured LTS, we instead employ the modal language used for expressing properties of an LTS (coloured or uncoloured) to represent a set of explicit norms, and then we use those norms to colour a regular (uncoloured) LTS. This is an advantage because it works with any transition system, not just those constructed using $n\mathcal{C}+$. It allows us to talk separately about the space of all possible non-normative behaviours — states, actions and their effects as represented by an uncoloured LTS — and norms regulating those behaviours, represented as a colouring of that LTS. Given an uncoloured transition system describing the possible non-normative behaviours, we have a notion of consequence between a set of explicit norms as established by some regulating authority and the obligations and permissions implied by those norms given that range of possible behaviours.

The second contribution is the treatment of connections between norms on states ('ought to be') and norms on transitions ('ought to do'). Instead of treating well-formedness constraints such as the 'green-green-green' constraint and its variants as fixed features of a coloured LTS we can instead define them within the modal language used for representing norms. That has the advantage of making their effects explicit and making it easy to add further optional forms in similar style.

We are currently investigating ways of extending and generalising the framework, in particular to allow defeasible conditional explicit norms with a priority structure to express which norms are intended to be exceptions to which.

Acknowledgements. This research was supported by the National Science Centre of Poland (grant UMO-2017/26/M/HS1/01092). We would like to thank the anonymous reviewer of the article for valuable suggestions, constructive criticism and encouragement.

References

[1] Åqvist, L., "Deontic logic", pages 147–264 in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, Second Edition, Vol. 8, Springer, 2002. DOI: 10.1007/978-94-010-0387-2_3

[2] Carmo, J., and A. J. I. Jones, "Deontic database constraints, violation and recovery", *Studia Logica*, 57, 1 (1996): 139–165. DOI: 10.1007/BF00370673

[3] Craven, R., and M. Sergot, "Agent strands in the action language nC+", *Journal of Applied Logic*, 6, 2 (2008): 172–191. DOI: 10.1016/j.jal.2007.06.007

[4] Giunchiglia, E., J. Lee, V. Lifschitz, N. McCain, and H. Turner, "Nonmonotonic causal theories", *Artificial Intelligence*, 153, 1–2 (2004): 49–104. DOI: 10.1016/j.artint.2002.12.001

[5] Hilpinen, R., and P. McNamara, "Deontic logic: A historical survey and introduction", pages 1–134, chapter 1, in D. Gabbay, J. Horty, R. van der Meyden, X. Parent, and L. van der Torre (eds.), *Handbook of Deontic Logic and Normative Systems*, Vol. 1, College Publications, London, 2013.

[6] Kulicki, P., and R. Trypuz, "Two faces of obligation", pages 70–82 in A. Brożek, J. Jadacki, and B. Zarnic (eds.), *Theory of Imperatives from Different Points of View (2)*, "Logic, Methodology and Philosophy of Science at Warsaw University 7", Wydawnictwo Naukowe Semper, 2013.

[7] Kulicki, P., and R. Trypuz, "Connecting actions and states in deontic logic", *Studia Logica*, 105, 5 (2017): 915–942. DOI: 10.1007/s11225-017-9716-1

[8] Meyer, J-J. Ch., "A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic", *Notre Dame Journal of Formal Logic*, 29, 1 (1988): 109–136. DOI: 10.1305/ndjfl/1093637776

[9] Sergot, M., "Modelling unreliable and untrustworthy agent behaviour", pages 161–178 in B. Dunin-Keplicz, A. Jankowski, A. Skowron, and M. Szczuka (eds.), *Monitoring, Security, and Rescue Techniques in Multiagent Systems*, Advances in Soft Computing, Springer-Verlag, 2005. DOI: 10.1007/3-540-32370-8_11

[10] Sergot, M., "Action and agency in norm-governed multi-agent systems", pages 1–54 in A. Artikis, G. M. P. O'Hare, K. Stathis, and G. Vouros (eds.), *Engineering Societies in the Agents World VIII. 8th Annual International Workshop, ESAW 2007, Athens, October 2007, Revised Selected Papers*, LNCS 4995, Springer, 2008. DOI: 10.1007/978-3-540-87654-0_1

[11] Sergot, M., "Some examples formulated in a 'seeing to it that' logic: Illustrations, observations, problems", pages 223–256 in T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*, Vol. 2 of "Outstanding Contributions to Logic", Springer Open, 2014.

[12] Sergot, M., and R. Craven, "The deontic component of action language nC+", pages 222–237 in L. Goble and J-J. Ch. Meyer (eds.), *Deontic Logic and Artificial Normative Systems. Proc. 8th International Workshop on Deontic Logic in Computer Science (DEON'06), Utrecht, July 2006*, LNAI 4048, Springer Verlag, 2006.

[13] Trypuz, R., and P. Kulicki, "On deontic action logics based on Boolean algebra", *Journal of Logic and Computation*, 25, 5 (2015): 1241–1260. DOI: 10.1093/logcom/ext057

[14] van der Meyden, R., "The dynamic logic of permission", *Journal of Logic and Computation*, 6, 3 (1996): 465–479. DOI: 10.1093/logcom/6.3.465

PIOTR KULICKI and ROBERT TRYPUZ
Faculty of Philosophy
The John Paul II Catholic University of Lublin, Poland
{kulicki, trypuz}@kul.pl
https://orcid.org/0000-0001-5413-3886
https://orcid.org/0000-0003-4042-9947

ROBERT CRAVEN and MAREK SERGOT
Department of Computing
Imperial College London
{robert.craven, m.sergot}@imperial.ac.uk
https://orcid.org/0000-0002-3706-1419