

Modelling Institutions using Dynamic Deontics

Christopher Frantz¹, Martin K. Purvis¹, Mariusz Nowostawski², and
Bastin Tony Roy Savarimuthu¹

¹ Department of Information Science, University of Otago, New Zealand

² Faculty of Computer Science and Media Technology, Gjøvik University College, Norway

Abstract. We have developed a refined institutional scheme derived from Crawford and Ostrom’s *Grammar of Institutions* (also referred to as ADICO) that has been adapted for the detailed representation of conventions, norms, and rules. In this work we apply this schema to model the emergence of norms. While previous work in the area of normative agent systems largely represents obligation and prohibition norms by discrete deontic primitives (e.g. ‘must’, ‘must not’, ‘may’), we propose the concept of dynamic deontics to represent a continuous perspective on emerging institutions. This supports the expression of norm salience based on the differentiated internal representation among participants.

To demonstrate how it can be operationalised for dynamic modelling of norms in artificial societies, we apply nADICO to a simple agent-based simulation. Our intention is to arrive at a dynamic modelling of institutions in general, facilitating a movement beyond the artificial boundaries between different institution types, while making the institutional grammar purposeful for a wide range of application domains.

Keywords: Dynamic Deontics, Institutions, Norms, Grammar of Institutions, Nested ADICO, nADICO, Reinforcement learning, Social learning, Norm enforcement, Multi-Agent Systems

1 Introduction

Crawford and Ostrom’s *Grammar of Institutions* [5] (GoI) is an approach to express social organisation (institutions) of different kinds, such as shared strategies (or *conventions*), *social norms*, and *codified rules*, using a unified grammar that not only integrates those different perspectives but supports the discrimination between those different institution types. To do so, the grammar consists of five components, *Attributes*, *Deontic*, *Aim*, *Conditions* and an *Or else* – ADICO in short – that are necessary to specify rules. By restricting constitutive components to a minimum, this syntax affords a wide scope for the expression of various institutional statements, such as norms and conventions, which we refer to as institution types for the remainder of the paper.

The generality of ADICO enables researchers to express various institutional views, including *institutions as equilibria* [15] (championed in the area of economic analysis), *institutions from a normative perspective* [28] (which concentrates on the behavioural perspective and is favoured by many researchers in the field of multi-agent systems (e.g. [2,29])), and *institutions as rules*, e.g. [17] (which is favoured by the New Institutional Economics movement [17,5]).

Previous approaches, such as Ghorbani et al.’s [8] work as well as Crawford and Ostrom’s conceptualisation, use the original grammar for a comprehensive description of *existing* institutions along with the social entities that shape and abide to them.

Notwithstanding the grammar’s attempt to represent institutions in a comprehensive manner, in this work we review central limitations of the grammar in its current state and discuss a refined formalisation of the grammar we have proposed in previous work [6]. In this paper we develop a more dynamic perspective on the grammar’s prescriptive component that is geared towards facilitating the bottom-up *emergence and establishment* of institutions we observe in human societies.

Accordingly the key element is our use of a *continuous notion of deontics* as an alternative to rigid deontic primitives, such as **must**, **must not**, and **may**, that are often associated with the use of deontic logic [31]. Our modification allows for less rigid and more fuzzy representations of institutions across individuals, while allowing the representation of fluid change over time. Along with this increased scope of expression, dynamic deontics can be used as an indicator of relevance offering the capability of weighing and prioritising potentially conflicting norms. This aspect allows for the modelling of the dynamic emergence of norms and their evolution over time, along with the representation of the important characteristic of stability that institutions exhibit.

In the next section (Section 2) we review Crawford and Ostrom’s grammar and its adoption in different fields. Then in Section 3 we present Nested ADICO (nADICO), which extends the feature set of the existing grammar and allows for a more detailed representation of institutions, including characteristics particular to institutions themselves (such as institutional regress). Following this, in Section 4, we introduce the notion of dynamic deontics that further refines the institutional grammar to enable the modelling of dynamic institutional environments. We demonstrate an executable agent-based model that uses the extended institutional grammar to dynamically generate institutional statements in Section 5. In Section 6 we summarise and contextualise this work and provide directions for future work.

2 The Institutional Grammar

2.1 Overview

The original ADICO grammar consists of five components. Those include:

- *Attributes* – describe the attributes and characteristics of social entities (which can be individuals or groups) that are subject to the institutional statement (e.g. convention, norm, rule). If not specified explicitly, all individuals (or members of a group/society) are implied.
- *Deontics* – a deontic primitive that describes either an *obligation* (e.g. represented as **must**), *permission* (**may**), or a *prohibition* (**must not**). In Crawford and Ostrom’s conception [5] it captures the aspects of deontic logic.
- *AIM* – describes an action or outcome associated with the institutional statement.
- *Conditions* – capture the circumstances under which the statement applies. This can include spatial, temporal, and procedural elements. If not further constrained, conditions default to “at all times and in all places” [5].

- *Or else* – describes consequences that are associated with the violation of the institutional statement, i.e. the combination of all other components used in that statement.

Crawford and Ostrom not only specify the components of the grammar; as indicated in the first section, the particular power of the grammar lies in its ability to satisfy different views on institutions, expressed as conventions, norms, and rules.

Using these three statement types, one can construct institutional rules of increasing prescriptiveness.

For a shared strategy (AIC), or convention, we can say:³

Drivers (A) hand their driver’s license to the police officer (I) when stopped in traffic control (C).

It effectively reflects a description of drivers’ commonly observable behaviour when facing the request to hand over their license. From a normative perspective, this can be interpreted as a descriptive norm.

In the GoI, a norm would extend a shared strategy with a prescription (and thus be equivalent to an injunctive norm), expressed as ADIC:

Drivers (A) must (D) hand their driver’s license to the police officer (I) when stopped in traffic control (C).

This represents an unambiguous instruction to the driver who (if taking a strictly deontological perspective) perceives it as his duty to present his driver’s license, independent of any threatening consequences.

Finally, a rule (ADICO) would introduce consequences for non-compliance:

Drivers (A) must (D) hand their driver’s license to the police officer (I) when stopped in traffic control (C), or else the police officer must enforce it based on traffic law (O).

Here the driver faces explicit consequences, which, depending on the nature of his refusal, can result in material (e.g. fines) or physical sanctions (e.g. arrest).

2.2 Application fields, Refinements and Limitations

The ADICO grammar provides a semi-formal description of operational institutional rules that make them accessible for institutional analysis [18] and structured policy coding [26]. In the area of multi-agent simulation, Smajgl et al. [27] have used the grammar to model endogenous changes of ADICO rule statements in the context of water usage. Significant recent contributions that use the grammar in more depth include Ghorbani et al.’s MAIA framework [8], which represents a comprehensive attempt to translate Ostrom’s Institutional Analysis and Development Framework [18] into an agent-based meta-model. Earlier, Ghorbani et al. [7] explored the notion of shared strategies as a fundamental statement type and differentiated their application across common, shared, and collective strategies.

³ In the following examples, we put in brackets the respective grammar component that represents the preceding fragment of the encoded institutional statement (e.g. ‘Drivers’ representing the *Attributes* component of the institutional statement).

Apart from a wide range of uses, the grammar has attracted some suggestions for refinement [22]. Our own work in this area is driven by the interest to make the grammar more flexible and dynamic. In this context we wish to highlight two key issues of concern, the first of which has been addressed and discussed in previous work [6].

First, the existing ADICO differentiation between shared strategies, norms, and rules (differing grammar components are used in those separate contexts) seems to limit the grammar’s ability to capture the notion of a norm in its full extent. In original ADICO terms, rules are assumed to have sanctions, whereas norms do not – at least not specified ones [5]. A further limitation is the lack of an ability to model the direct dependency of institutions in a specified systematic manner, i.e. the rules another rule depends on for its enforcement, such as ‘sanctioning the sanctioners’ in case of non-compliance, which we think is crucial to provide an authentic representation of codified rules/formal institutions in particular, but also offers alternative means to differentiate norms and rules (see Subsection 3.2).

Second, in ADICO the notions of prohibition and obligation norms are mapped into a “boolean” [9] perspective. Other authors have already pointed out this limitation and argued for a more continuous perspective, both for the ADICO grammar [22] and for social norms in general [9]. Particularly when conceiving institutions as emergent properties of societies (as opposed to intentionally constructed), modelling the progression from individual behaviour to social behaviour across differing institutional types requires more flexibility in specifying norms, beyond the discrete **mays**, **musts**, and **must nots**; if not prescribed by some social authority (e.g. leader), the rigid prescriptive deontics are an unlikely starting point of institutional development. In practice, more flexible boundaries are desirable to support continuous adaptation so that a new and different norm may gradually emerge from an existing one or simply gain salience and replace a norm that reached the end of its life cycle. Given the interpretation of norms as implicitly shared representations, they are subject to subjective perception and evaluation by norm participants, an aspect we can observe in the daily use of language (e.g. use of ‘should’ instead of ‘must’). In that context particularly the permissive primitive **may** is of limited value when describing behavioural regularities. Apart from constituting the right to take an action [5,24], its concrete meaning relies on internal individual utility evaluations (e.g. Crawford and Ostrom’s deltas) and is often insufficient to express observable social norms.

Attributing stronger descriptive power of norms by offering a more fluid representation is in line with demands raised by institutional scholars [12]. The selective use of Dynamic Deontics emphasises the general nature of the institutional grammar, beyond the refinements offered by Nested ADICO.

3 Nested ADICO (nADICO)

This section provides a brief overview of the refinements suggested as part of Nested ADICO (nADICO). Earlier [6], we discussed those refinements in more detail.

To address the limited expressiveness and unbalanced representation of different institution types in the GoI, we refine the GoI by introducing the following three central amendments:

- Representation of Sanctions for Norms
- Systematic Nesting of Institutional Statements
- Refined Differentiation between Norms and Rules

3.1 Nesting Capabilities

Crawford and Ostrom’s grammar uses the ‘Or else’ component, which is used to express sanctions, including a notion of nesting of institutional statements. However, the unstructured manner of the sanction component limits the computational representation, but also does not exploit the grammar’s potential. To extend the comprehensiveness of the grammar (in particular with respect to norms), opening it for a more dynamic perspective and improving its computational accessibility, we can back institutional statements with statements that bear the same structural components (an aspect that was considered by Crawford and Ostrom [5], but not systematically explored).⁴ This entails developing a nesting structure of institutional statements consisting of the ADIC components of the original GoI. Extending the example from Section 2, we can thus capture consequences associated with a given rule breach, and do so for an arbitrary number of nesting levels, reflecting the notion of institutional regress.

Drivers (A_1) must (D_1) hand their driver’s license to the police officer (I_1) when stopped in traffic control (C_1),
OR ELSE ————— *2nd level* —————
the police officer (A_2) must (D_2) enforce this (I_2) under any circumstances (C_2),
OR ELSE ————— *3rd level* —————
internal investigators (A_3) must (D_3) follow up on this issue (I_3) in any case (C_3).

Vertical Nesting – Given the introduction of different levels that are activated upon institution violation on the preceding level, we call this nesting type *vertical nesting*. Using the grammar primitives, we can express this structure as $ADIC(ADIC(ADIC))$, where the respective leading statement represents the *monitored statement* (obligation of drivers to hand over license) that activates a *consequential statement* (police officer’s obligation to enforce it). In this case, the ‘drivers’ (A_1) are potential first-order violators, and the police officer is first-order sanctioner (A_2). Looking at extended nesting levels, the police officer is likewise a potential second-order violator, and internal investigators (A_3) represent second-order sanctioners, and so on. In equivalence, the first-order consequential statement is likewise second-order monitored statement, etc.

Horizontal Nesting – Apart from facilitating the representation of institutional regress, nADICO further introduces a notion of *horizontal nesting*. The purpose of this is to provide more detailed modelling capabilities by avoiding a strict 1:1 assignment of monitored and consequential statements. One can imagine a variety of different gradual sanctions imposed upon the individual (e.g. speeding may result in an instant fine as well as an increment in demerit points), which may be applied in conjunction or alternatively. Especially for the normative case, in which consequences may not be formally

⁴ In the context of normative multi-agent systems, the nesting of norms based on their function (e.g. substantive norm backed by check norm) has been discussed by Grossi et al. [10].

specified, sanctions can be unpredictable, e.g. sanctions for observing a jaywalker may extend from scolding to physical abuse, or none may be applied. For this purpose, nADICO introduces different logical operators that allow the expression of statement combinations. The operators include logical conjunction (*and*), inclusive disjunction (*or*) and exclusive disjunction (*xor*). Their use allows the expansion of simple ADIC statements into statement combinations such as (ADIC *and* ADIC) on a given level, which could likewise be nested (e.g. (ADIC *and* (ADIC *xor* ADIC))) to express complex institutional constructs.

Expanding the previous example into the structure ADIC((ADIC *and* (ADIC *xor* ADIC))ADIC), we could express:⁵

Drivers (A_1) must (D_1) hand their driver's license to the police officer (I_1) when stopped in traffic control (C_1),

OR ELSE ————— *2nd level* —————

the police officer ($A_{2a/b/c}$) must (D_{2a}) enforce this (I_{2a}) under any circumstances (C_{2a})

and,

depending on severity ($C_{2b/c}$), must ($D_{2b/c}$)

either fine the driver (I_{2b})

or arrest him (I_{2c}),

OR ELSE ————— *3rd level* —————

internal investigators (A_3) must (D_3) follow up on this issue (I_3) in any case (C_3).

Note that horizontal nesting can likewise occur in monitored and consequential statements (which is consequent, knowing that monitored statements can be consequential statements with respect to different institutional statement levels).

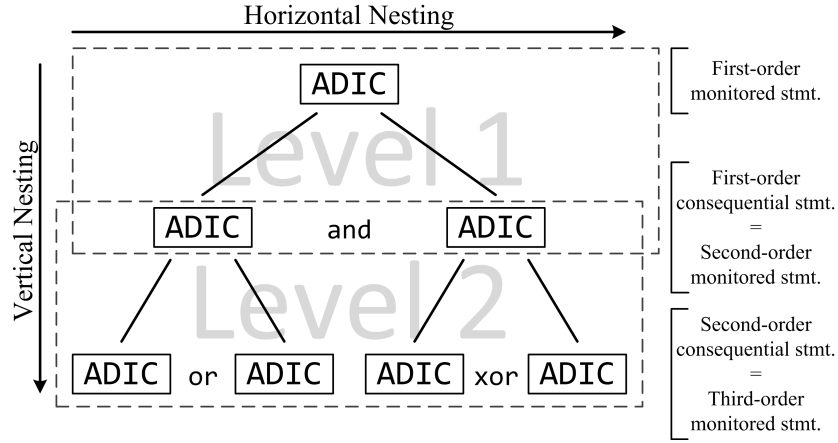
Figure 1 visualises the nesting capabilities of nADICO in an exemplified manner. An extended description of nADICO along with its formalisation can be found in [6].

3.2 Refined Differentiation between Norms and Rules

The introduction of sanction specifications for norms requires a revised grammar interpretation, as we lose the ability to syntactically differentiate norms and rules purely based on the existence of sanctions. However, by introducing nested institutional statements we gain the ability to inspect the characteristics of respective nested statements. A characteristic of norms is their generally distributed enforcement and the associated nature of the norm monitor (and potential sanctioner), an entity that is not represented in the original GoI. Monitors can be internal (e.g. unconscious self-monitoring), self-assigned or informally assigned, and beyond that, particularly for less salient norms, it can be hard to know who monitors the given norm after all, and if so, what the nature of the sanctions associated with a violation is. Expressing norms in nADICO, we would expect a fuzzy representation of the sanctioner and likewise sanctions. To express the

⁵ Note that we extend the index indicating the nesting levels along with letters that associate grammar components with the respective consequential statement(s) on that level. In this example, the second level comprises three statements (a, b and c), all of which share a common sanctioner A_2 , expressed as $A_{2a/b/c}$, but only b and c share the same *Conditions* ($C_{2b/c}$) and so on.

Fig. 1: Nesting Characteristics of nADICO



varying application of sanctions, the introduced logical operators can facilitate the differentiation between different institution types, with the ‘inclusive or’ (*or*) implying some fuzziness as to which one(s) and how many concrete sanctions out of a selection may be applicable in a given situation.

On the other hand, the existence of a well-specified formally assigned monitor is a characteristic for rules.⁶ Given that the specification clarity is a key differentiation criterion between norms and rules, the nature of rules can be further associated with the use of *and* and *xor* operators if horizontal nesting is applied, inasmuch as those allow an unambiguous specification of sanctions in contrast to *or*.

An important aspect the original GoI does not consider is the *differentiation between rule monitor and sanctioner/enforcer*. In nADICO we introduce not only the clear specification of the sanctioner, but, where applicable and possible, the explicit specification of sanctioner and monitor. From an operational perspective, this can, again, be facilitated using horizontal nesting of statements that allow the specification of duties for both the monitor and sanctioner on a given nesting level.

Table 1 summarizes the discussed differentiation mechanisms. Note that the differentiation highlighted here smoothens the crisp boundaries between norms and rules and may not capture all imaginable cases, but offers a more detailed and realistic encoding of institutional complexity.

4 Dynamic Deontics

4.1 Concept and Characteristics

As mentioned in Section 2.2, the restriction to three deontic primitives is not sufficient to represent the mechanisms by which institutions evolve and the way they change over time. This rigidity is primarily due to the discrete primitives **must** and **must not** that

⁶ For a detailed overview of potential monitor types and their association to institution types, refer to [6]. Also note that, beyond the differentiation mechanisms discussed here, and in line with the original ADICO grammar, nADICO relies on the meta-norm of collective action to constitute rules.

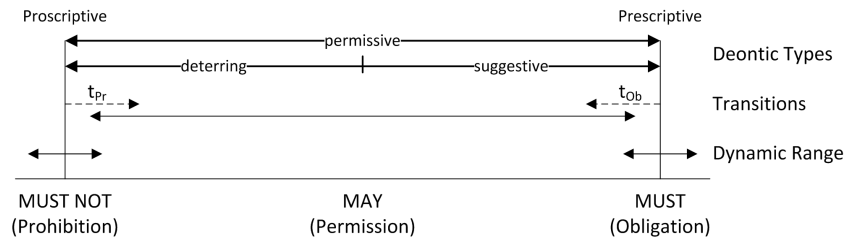
Table 1: Differentiating Characteristics for Norms and Rules in nADICO

Characteristic	Norms	Rules
Specification of Monitor	unspecified / fuzzy	clear specification
Specification of Sanctioner	unspecified / fuzzy	clear specification
Assignment of Monitor / Sanctioner	informal	formal
Relationship between Monitor / Sanctioner	often unified entity, not explicitly specified	unified or separated, clear specification
Nature of Monitor	see Monitor Types specified in [6]	
Combination Operators	<i>or</i>	<i>and, xor</i>

represent obligations and prohibitions, which reflect commonly accepted notions of social norms [23,9], particularly in the context of normative multi-agent systems (e.g. [2]). In contrast to those two strict injunctions stands the permissive primitive **may**, which remains imprecise concerning its associated duties.⁷ We introduce three aspects that are central to a more continuous notion of deontics, before discussing underlying conceptual implications.

Continuous Notion of Deontics – Instead of relying on a strict tripartite structure of deontics, we believe that a more straightforward way to deal with this consistency issue is to allocate deontic values on a continuous scale (an aspect von Wright [30] was already aware of), delimited at the extremes by a prescription (obligation, i.e. **must**) and a proscription (prohibition, i.e. **must not**) advocating a gradual understanding of norms, a schematic visualization of which with respect to an aim (i.e. an action or an outcome in the sense of the ADICO syntax) is provided in Figure 2. At the extremes we allocate **must** and **must not**, with more permissive points in between, effectively capturing the *omissible* and *promissible* to a varying extent⁸. This approach underlies the assumption that an institutional statement is associated with a valence that drives it either towards prescription or proscription, irrespective of whether it ever reaches one of those two extremes.

Fig. 2: Dynamic Deontic Scale



Stability – Using this continuous-scale perspective depicted in Figure 2, we can model institutional emergence and also identify the relative importance of various institutions.

⁷ These incongruencies of the deontics were recognised by Crawford and Ostrom [5], in particular the contrast between the semantics of permission and those of a prescription of duty [25].

⁸ The *omissible* describes an obligation from which we can deviate in exceptional cases; under *promissible* we understand a prohibition which we can exceptionally deviate from.

In addition, we can use this scheme to represent stability, a key aspect of prohibition and obligation norms. Norms that have reached the extremes of the normative scale tend to show strong change resistance – thus once settled on a **must** or **must not** (e.g. prohibition of homosexuality), they often become stubbornly entrenched. We can represent this ‘stickiness’ of prohibition and obligation statements by introducing tolerance regions around the deontic extremes, denoted by t_{Pr} and t_{Ob} in Figure 2, which are associated with conditions that prevent the rapid change of extreme deontic values. One could likewise introduce a notion of friction or viscosity to constrain the movement and thus have uniform or differing stability characteristics along the deontic scale.

Dynamic Deontic Range – Measures of extremes are taken from an individual utilitarian viewpoint. The width of the range is thus based on the personal experience of the agent along with preimposed moral dictate based on family, culture or religion. With experience, one’s moral views evolve, and his or her subjective range between **musts**, **may**s, and **must not**s will be adjusted. A relatively inexperienced agent may have a narrow deontic range and have many attitudes lodged at the extreme positions that have been imposed, e.g. by preimposed religious beliefs. But as one is exposed to a wider range of experiences (e.g. in the case for attitudes concerning homosexuality, one is exposed to a wider range of views and backgrounds on this issue), adjusted experiences may lead to an expanded or more nuanced deontic scale that captures a more refined viewpoint. We suggest that this dynamic deontic scale can expand and contract throughout an individual’s lifetime, both based on reinforced or subsiding external stimuli as well as adopted viewpoints.

4.2 Discussion

At this stage, we wish to elaborate on the philosophical underpinnings that motivate dynamic deontics⁹. When analysing norms in a given society, those are conventionally assumed stable [23] and objectified using a unified representation (e.g. All agents think: ‘An agent must not cheat’) that allows their explicit sharing and unambiguous understanding. Utilizing a varying degree of salience for different norms [4], individuals then decide whether to comply with such norms depending on their situational disposition. In this view the assumption of a unified understanding is a generally accepted convenient modelling abstraction, but it does not take into account the different nature of individuals based on their background¹⁰ and experience, an aspect the concept proposed here captures by incorporating a dynamic deontic range. By respecting the fluidity and nuanced nature of norms, we can leverage norms as artifacts that describe the society they act in.

The concept of Dynamic Deontics adopts this perspective and does not assume (but permits) the explicit communication of norms but allows the development of subjective norm understanding based on experience (see e.g. Savarimuthu et al.’s work on norm

⁹ Note that the concept of dynamic deontics is not to be confused with dynamic deontic logic [16], which formalises norms over action as opposed to norms over states, an aspect pointed out by one of the anonymous reviewers.

¹⁰ Henrich [11] describes the varying emotions associated with the perception of obligation and prohibition in different cultures.

learning [21]) that relaxes the assumption of a unified norm understanding (e.g. Agent 1 thinks: ‘An agent *must not* cheat.’; Agent 2 thinks: ‘An agent *should not* cheat.’), which, in the light of differing exposure of individuals in open societies, hardly seems realistic. Instead we can conceive norms as inherently distributed in their nature, which includes the subjective understanding within individuals. Accepting the individualised understanding of norms by assigning varying deontics along a deontic range, the norm in the society can then be described as the aggregate of the individual norm perceptions, i.e. the collective understanding of what the norm is. The level of agreement on that norm within a society then is a property of that particular norm instance (i.e. the norm in the context of the society it acts in) itself, which, over longer time frames, allows the representation of fluidity of social norms.

It is important to note that the individualised understanding of norms is not to be confused with the individuals’ attitudes towards that norm. Individuals do not autonomously align norms and their attitudes towards them unless they have the power to do so (e.g. by influencing others); instead the individualised norm understanding (and in principle also representation) is subject to modification based on social norm transmission processes (e.g. norm enforcement, social learning) individuals are exposed to. In addition to the “boolean” mode of the deontic primitives, the modelling of norms in a more fluid (or viscous) nature, as demanded by various institutional scholars [12], makes a norm’s specific nature (including aspects such as the aggregate understanding across as well as diverging understanding within a society) a characteristic of the society it describes.

5 Simulation

5.1 Model

Using the nADICO grammar, we introduce an agent-based simulation model that demonstrates how agents could leverage some of the nADICO characteristics and dynamic deontics, in coordination with reinforcement learning (RL) [32] based on the reward from the environment and social learning [3] (i.e. based on the rewards obtained from others). Doing so, we take a consequentialist perspective on norms, thus suggesting the adoption of norms based on experience, as opposed to the conventionally assumed deontological ethics perspective in which individuals act with respect to a known ‘Right’ (as opposed to ‘Wrong’), which requires the existence of preimposed norms. However, the operationalisation proposed here assumes a greenfield approach without pre-existing norms.

In this experiment each agent in the model has a set of actions it can perform (its *action pool*) as well as a set of reactions to other agents’ actions (rewards/punishments – *reaction pool*). The operation of the model employs a simple trade metaphor. Imagine that two agents, A and B, get together for a transaction. A can hire B to sell some of his goods for a commission. B in this scenario has two options: it can return the cash it received to A (*trade fairly*) or cheat A out of the money (*withhold profit*). In response to either of these actions, A can choose one of the following rewards/punishments (reactions) for B: a) Pay the commission to B, b) Refuse to pay commission to B, c) Fire

(dismiss) B from further employment, d) Retaliate against B’s family. The respective utilities for these actions and reactions are given in Table 2.¹¹

Table 2: Action Sanction Feedback

Action-Reaction Combinations		Utility from Actions	
Action (Agent B)	Reaction (Agent A)	for Agent B	for Agent A
TRADE FAIR	DON'T PAY COMMISSION	-0.5	1
TRADE FAIR	FIRE	-1	-0.5
TRADE FAIR	RETALIATE FAMILY	-1.5	-0.5
TRADE FAIR	PAY COMMISSION	0.5	0.5
WITHHOLD PROFIT	DON'T PAY COMMISSION	0.5	-0.5
WITHHOLD PROFIT	FIRE	-0.5	0
WITHHOLD PROFIT	RETALIATE FAMILY	-1.5	1
WITHHOLD PROFIT	PAY COMMISSION	1	-1

Each agent maintains its own memory instance (here: Q-Learning). Since the assembled agents can act as both actors and reactors, the utility response from actions is fed into a combined RL instance structure, which is used for both the choice of actions and reactions. Since agents don’t have intrinsic knowledge about the value of their actions *ex ante*, the model uses RL and/or social learning to build social norms to guide behaviour.

Agent Strategies – At the beginning of each simulation round, based on the exploration probability, agents choose to

- *exploit* (engage directly in trade by interacting with another agent), or
- *explore* (learn more about its environment for future benefit by observing other agents).

When an agent is in *exploitation* mode, it chooses reward-maximizing actions and reactions based on what it has learned from past experiences. In this mode the agent can also be assigned by the modeller to engage in third-party norm enforcement. In this case the third-party agent observes the action of another agent (which is involved in an interaction with some other agent), and it carries out its own reaction (reward or punishment) on this observed agent, irrespective of whatever reaction the observed agent received from its own trading partner. To facilitate this operation, the modelling environment has each agent display its most recent action and the reaction it received, along with how the agent considered that received reaction (expressed as a valence of -1, 0, or +1).

When an agent is in *exploration mode*, the agent randomly chooses an action to see if it “works”. The reaction it receives from this randomly chosen action will be remembered for future learning purposes. In this exploration mode, the agent can also observe action-reaction activities from randomly chosen other agents and thereby learn about consequences. Actions are expressed as ADIC statements, with the deontic component (D) set to zero, which implies a neutral perception towards the action (which is the center of the deontic scale in Figure 2). In terms of the deontic triad, we can interpret this as a **may**. Action-reaction combinations are then ADIC(ADIC). Observing agents use the valence (-1, 0, or +1) associated with the visible action-reaction combination to

¹¹ To facilitate the interpretation, we only introduce two actions for this simulation.

approximate what the observed agent received as a reaction and thereby learn from the observation.

Operationalising Dynamic Deontics with Reinforcement Learning – Although in a fixed social world the deontic value range could be held static, we are interested in environments where deontic values can change over time. Under these circumstances each agent maintains min and max values that may vary. We offer a sample semantic mapping: at the extremes (i.e. at min. and max.) we associated **must** (max.) and **must not** (min.); between those extreme values we associate the values **should**, **may**, **may not**, and **should not** (which could be allocated along the scale in Figure 2). For this initial operationalisation we put emphasis on simplicity and assume the compartments of the respective deontics to be of equal size and symmetric along the deontic scale.

In the present context, the min and max values are based on the agent’s Q-values stored in its memory. Using a sliding window approach, the mean of a fixed-length history of highest Q-value defines the prescriptive end; likewise the mean of the lowest Q-value history specifies the proscriptive boundary of the deontic range. In this case, RL operationalises both the expansion as well as the reduction of the deontic range (based on the reinforcement and discounting of Q-values at the end of each round). The Q-values are collected for action-reaction combinations, with experienced consequences combined using the *or* operator¹². Transforming RL memory entries to institutional statements, action-reaction sequences are aggregated by action using nADICO’s horizontal nesting capabilities (see Section 3). Let us assume, that $stmt_l$ indicates a statement on l^{th} level, and $stmt_{l+1,i}$ indicates i^{th} statement on $(l+1)^{th}$ level, $count_l$ indicates the number of statements on l^{th} level, and $c_{deonticRange}$ is the center of the deontic range. The deontic of the leading monitored statement $stmt_l$ ($d(stmt_l)$) is then derived by aggregating the consequential statements and depends on the logical connection of the consequential statements on a given nesting level, i.e. all the $stmt_{l+1,i}$ statements.

For *or* and *wor* combinations, the monitored deontic is the value of the consequential statement whose individual deontic shows the greatest deviation (extremal) from the center of the deontic range towards the direction indicated by the sum of all consequential deontics (*deontic bias*), i.e.

$$extremeDeontic(stmt_l) := \begin{cases} \max(d(stmt_{(l+1),0}), \dots, d(stmt_{(l+1),count_{(l+1)}})), & \text{if } (\sum_{i=0}^{count_{(l+1)}} d(stmt_{(l+1),i})) \\ > c_{deonticRange} \\ \min(d(stmt_{(l+1),0}), \dots, d(stmt_{(l+1),count_{(l+1)}})), & \text{otherwise} \end{cases}$$

However, the extreme deontic is only applied if the sum of the consequential deontics is not located at the deontic range center $c_{deonticRange}$, in which case the deontics of the nested statements cancel each other. In that case, the deontic range center itself describes the statement’s deontic (which, under the assumption of a symmetric deontic range, resolves to **may**), i.e.

$$d(stmt_l) := \begin{cases} c_{deonticRange}, & \text{if } (\sum_{i=0}^{count_{(l+1)}} d(stmt_{(l+1),i})) = c_{deonticRange} \\ extremeDeontic(stmt_l), & \text{otherwise} \end{cases}$$

¹² Note that this simulation takes the greenfield approach, i.e. individuals do not know about action effects. No collective action process takes place; no rules are specified ex ante.

Reason for choosing the extreme deontic is the assumed application of only one sanction at a time. The modelled agents are modelled as pessimistic and expect the most extreme individual sanction for a given action when interpreting the action.

For *and* combinations, the monitored deontic is the sum of all the consequential statements' deontic values as agents can assume the co-occurrence of sanctions combined by *and* operators, i.e.

$$d(stmt_t) := \left(\sum_{i=0}^{count_{(t+1)}} d(stmt_{(t+1),i}) \right)$$

The resulting deontic value for the aggregated action is then the agent's valuation of this action (irrespective of a potential reaction) in terms of its own system. Since the Q-value reflects both, qualitative feedback and frequency (i.e. probability of occurrence), the use of the maximum Q-value is well-suited for this purpose.

In an effort to reflect the experience that social norms (particularly at the extremal ends of the scale) tend to be enduring, even after they no longer reflect their original purposes, we have also incorporated a 'stickiness' mechanism in our simulation to be associated with the extremal (min and max) ends of the deontic scale. Thus associated with those two ends of the deontic scale are occurrence thresholds that determine whether an institution is worthy of being designated as an obligation or a prohibition norm based on the deontic range allocation. We operationalise this by tracking the number of rounds a deontic value reaches into the tolerance zone around an extreme deontic (t_{Pr} and t_{Ob}). This transition is parameterised using a stability threshold ($th_{establish}$) as well as a destruction threshold ($th_{destruct}$).

Algorithm 1 summarises the agents' execution cycle. The parameter set of the model is presented in Table 3.

Algorithm 1: Agent Execution Cycle

```

Pick two random other agents;
Decide whether to explore or exploit in this round;
if exploring then
    Pick random action from action pool;
    if social learning activated then
        Observe action of first randomly chosen agent and internalize action-sanction
        combination along with valence (not actual reward value);
    else
        Pick action with highest Q-value from action pool;
        if norm enforcement activated then
            Sanction action taken by first randomly chosen agent using sanction with highest
            Q-value;
            Memorize feedback from sanction choice;
    end
Execute picked action and apply to second randomly chosen agent;
Memorize reaction and make action-reaction combination (with valence representation of
feedback) visible to other agents;
Update deontic range;
Check for stability of nADICO sequences as well as shifts from/to obligation or
prohibition norms;

```

Table 3: Simulation Parameters

Parameter	Value
Number of agents	100
Tolerance zone around extreme deontics (t_{Pr} , t_{Ob})	0.1 of deontic range amplitude
Norm stability threshold ($th_{establish}$)	100 rounds
Norm destruction threshold ($th_{destruct}$)	200 rounds
Deontic range history length	100 rounds
Memory discount factor	0.99
Exploration probability	0.4

5.2 Results

The simulation runs comprised the following four configurations based on the combination of different social actions:

- No social learning, no norm enforcement (Scenario 1)
- Social learning, no norm enforcement (Scenario 2)
- No social learning, norm enforcement (Scenario 3)
- Social learning, norm enforcement (Scenario 4)

In all cases, agents receive direct feedback for their actions and likewise sanction others (positively or negatively) for actions imposed on them. In some simulation configurations, secondary indirect social interactions – norm internalising (social learning) and socialisation (norm enforcement) – were included. The simulation environment is based on our own simulation platform that uses the Mason simulation toolkit [13] for scheduling and visualisation support. We repeatedly ran each simulation configuration 30 times for 20,000 rounds to validate the outcomes, but given the explorative nature of this simulation, we describe the outcome of a representative simulation run. In the results shown in the following figures, we present the learned behaviour from the perspective of the role of Agent B (the hired trader). However, during the course of the simulation, agents can take both roles repeatedly and integrate their normative understanding towards those actions from both perspectives.

In the scenario that avoids any indirect social action (Scenario 1; Figure 3), all indirect social actions were excluded.¹³ Throughout these simulation runs, agents gradually developed the “understanding” that it is most beneficial to cheat (i.e. **must not** trade fair), a tendency that reached to 70-80 percent of the agents during the execution. There was a complementary, declining portion of agents that thought they **must** trade fairly. A less visible norm that arose included the suggestion that agents **may** withhold profit, while a significant number of agents maintain the understanding that they **may not**, and to a lesser extent, **should not** withhold profit. Overall, the graph in Figure 3 shows a diversity of views in the community with tendency to non-cooperation.

For Scenario 2 (Figure 4), for which social learning was incorporated, we can observe its significant effect on behaviour. Agents “mimic” other agents’ behaviours, and given that unfair trading dominates in the previous scenario that does not employ any social learning (Scenario 1), the performance of the community converges towards clear and extreme norms. The perception of unfair trading (**must not** trade fair) ranges at

¹³ Each figure consists of two subfigures (one for each action), visualising the distribution of the different deontic terms towards that action in the agent society.

around 100 percent; Agents increasingly think they **may** withhold profit (reaching 40-50 percent), complemented by gently declining percentages of agents feeling merely that one **may not** and (on a lower level) **should not** trade fair. The benefit here from the combined use of RL and social learning is compatible with previous findings (see e.g. [20]).

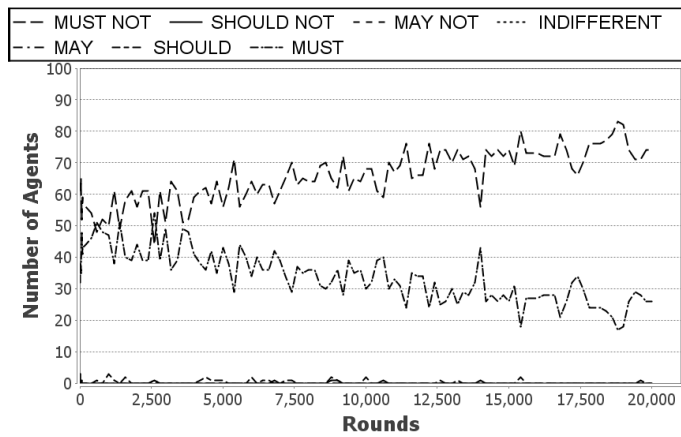
For Scenario 3 (Figure 5), which incorporated indirect norm enforcement by third parties but not social learning, an entirely different pattern from that of Scenario 2 emerged. Given that any additional social reaction here is based on previous actions on the part of an observed agent, norm enforcers act from the perspective of a hiring agent, thus rewarding fair trading and punishing unfair trading. As a result the obligation norm of fair trading (**must**) dominates and is increasingly supported by the complementary understanding not to withhold profits (**should not** withhold profit' at around 90 percent); less than 10 percent believe they **must not** withhold profit. At this stage it is important to note that although both mentioned actions ('trade fair', 'withhold profit') are seemingly complementary, their reinforcement (both positive and negative experiences) depends on the agents' situational choice, which during the course of the simulation (driven by the pay-offs defined in Table 2 and the fact that agents integrate the experience from both perspectives, both as acting (hired) agent and reacting agent) drives towards the dominant choice of the action 'trade fair'. Consequently, the normative reinforcement of this action exceeds that of 'withhold profit'.

The final scenario (Scenario 4; Figure 6) explores the combined use of social learning and norm enforcement. The outcome here enhances the effect of norm enforcement shown in Scenario 3, but improves the convergence by incorporating social learning effects. As a result, agents develop more extreme normative understandings (all agents believe they **must** trade fair; the number of agents thinking they **should not** withhold profit is increasingly replaced by the extreme understanding that withholding profit is prohibited (**must not**), reaching up to 20 percent at the end of the simulation run).

In addition to the macro view demonstrated by these simulations, it is useful to look at the individual agents' evolving understanding of norms (the emerging nADICO sequences and the situational deontic range). This enables one to see how Q-values are translated from this reinforcement learning-based approach into social consequences mapped to nADICO statements. The situational extract shown in Figure 7 taken from an individual agent of Scenario 4 shows how the agent develops the perception that it must trade fair (refer here to nADICO "Statement 1" in Figure 7), mostly driven by the threat of not being paid its wage (commission), which is shown here to be the most extreme deontic value in the agent's deontic range. Recall that the ADICO syntax is constructed to sanction nonadherence to monitored statements by threatening with an 'Or else'. However, the deontic values, here derived from Q-values, imply a "because" or "on the grounds that" relationship (e.g. 'I *must* trade fairly, *because* my employer *must* pay me my commission.'). In order to establish this semantic translation from Q-values to subjectively meaningful consequential statements, which includes a shift in perspective from subject to sanctioner, we invert the deontic associated with the particular Q-value. Effectively, the agent is using its own experience (Q-values) to engage in empathetic perspective-taking of the other observed agent, and anticipates what it might do as a reaction to the evaluated action. Thus the agents might surmise, 'I must trade

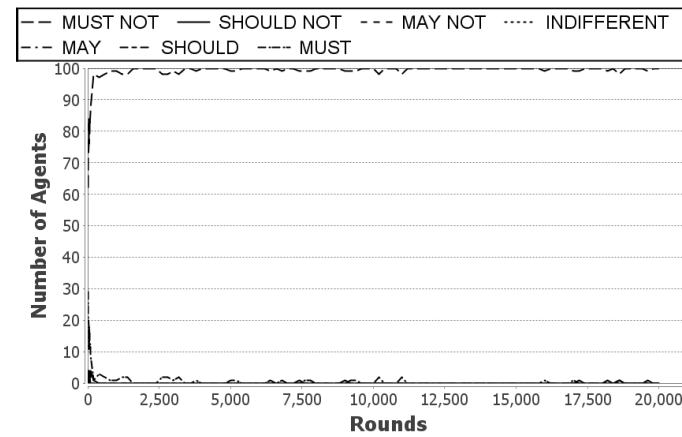
fairly, *or else* my employer *must not* pay me my commission.' In order to carry out that conjecture, we need to *invert* the deontic associated with a particular value in order to place it in the context of the other agent, which implies a shift from **must** to **must not**, **should** to **should not** and so on. Consequently, the agent bases its understanding on the negative consequences of being fired, not being paid wage, and family retaliation.

Fig. 3: Emerging Norms in Scenario 1
(No social learning, No norm enforcement)

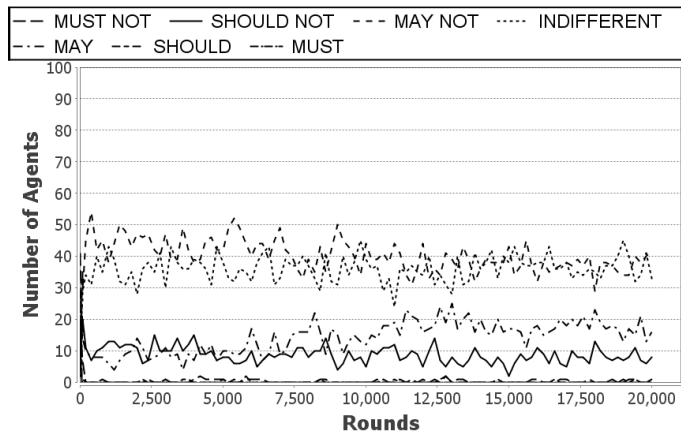


(a) Action 'trade fair'

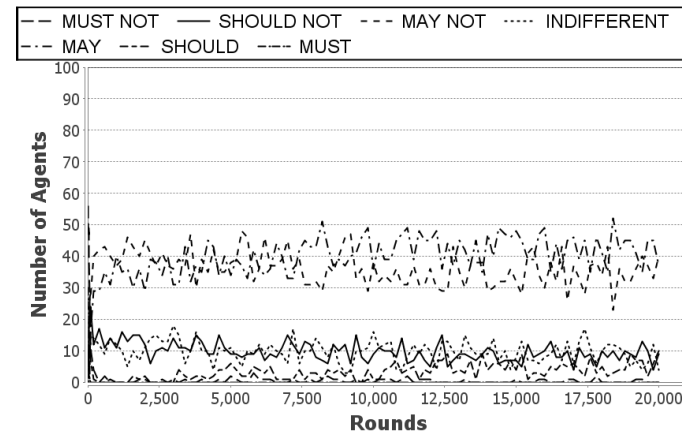
Fig. 4: Emerging Norms in Scenario 2
(Social learning, No norm enforcement)



(a) Action 'trade fair'

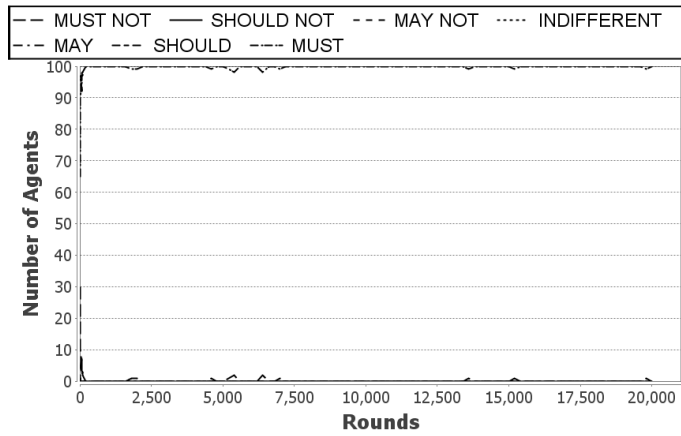


(b) Action 'withhold profit'



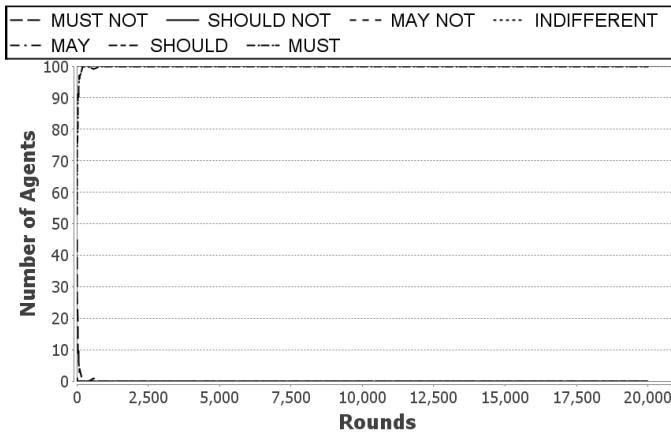
(b) Action 'withhold profit'

Fig. 5: Emerging Norms in Scenario 3
(No social learning, Norm enforcement)

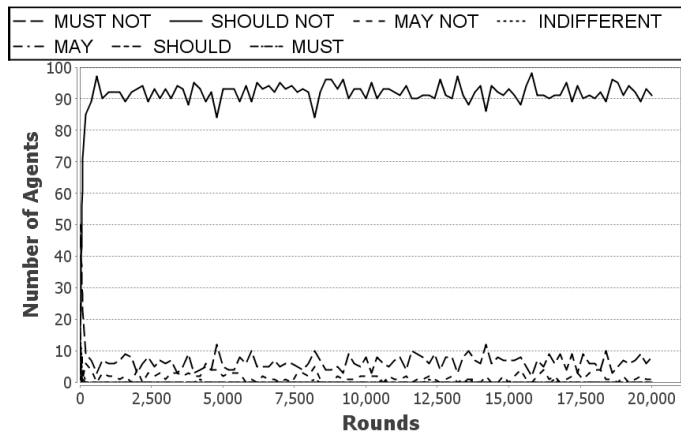


(a) Action 'trade fair'

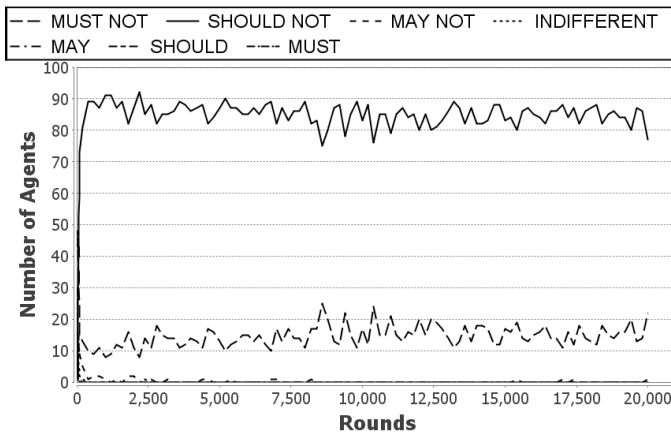
Fig. 6: Emerging Norms in Scenario 4
(Social learning, Norm enforcement)



(a) Action 'trade fair'



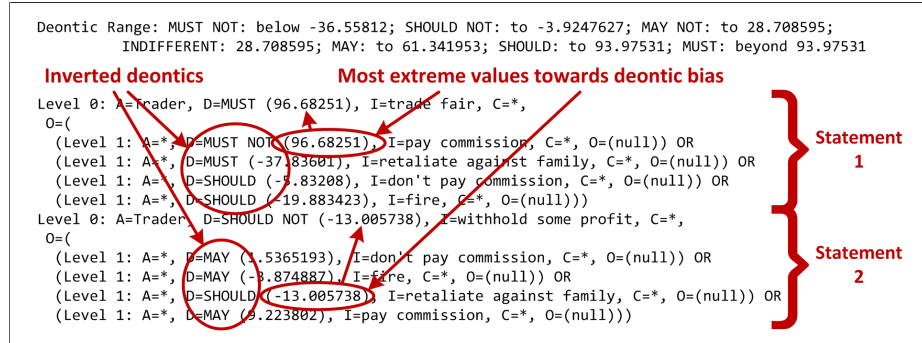
(b) Action 'withhold profit'



(b) Action 'withhold profit'

“Statement 2” in Figure 7 indicates that the trader should not withhold profit, or else retaliation against family may ensue as well as the other consequences. This example highlights the differentiated perceived threats when mapped to human-readable deontics. Note in this figure that while retaliation appears to be a dominating sanction (‘**should**’), other sanctions are associated with weaker prescriptions (‘**mays**’).

Fig. 7: Situational Deontic Range and generated nADICO Statements



6 Discussion and Future work

This paper discusses the introduction of dynamic deontics into nADICO, a refined ‘Grammar of Institutions’, with the intent to extend its capabilities to express the dynamic aspects of institutions, such as their emergence and change over time (continuity of deontics) as well as stability (establishment/destruction thresholds), while reflecting individual participants’ differentiated understanding of institutions (individual dynamic deontic ranges). We operationalised nADICO with dynamic deontics to model the establishment and change of norm understanding over time based on different scenarios that incorporated reinforcement learning, social learning, and norm enforcement as mechanisms to socialise norm understanding. The experiments described in this work take a greenfield approach and trade the strictly deontological perspective on social norms for a consequentialist perspective, in which agents develop an understanding of conduct by individual learning, observation, but also via norm enforcement by others, instead of relying on preimposed norms to regulate their behaviour. Existing institutional environments characterised by preimposed norms and rules can be represented by modification of the ‘stickiness’ behaviour but also the specification of predefined deontic ranges (both described in Section 4), an aspect left for future work. Particularly the ‘stickiness’ aspect is important to simulate the ‘lock-in’ effect of norms, i.e. the adoption and persistence of suboptimal norms, but also to model conflicting behaviour in culturally diverse environments.

In the area of normative multi-agent systems we can find a variety of approaches to model norms im/emergence and sanction-based enforcement (e.g. Andrighetto et al. [2]). Our simulation exemplifies an approach of dynamic sanctioning, but unlike Mahmoud et al.’s approach [14], it does not base its reaction on individual norm-

compliance but only on aggregate experience. Villatoro et al. [29] propose a more complex model of dynamically adjusted sanctioning based on a heuristic that, in addition to violation behaviour, incorporates sanctioning cost. Reflecting on work produced in the context of this volume, our approach shares intentions with and complements other contributions, such as Panagiotidi et al.'s [19] attempt to bridge the gap between norm formalisation and operationalisation as well as Aldewereld et al.'s [1] conceptualisation of group norms whose elements could be expressed using the nADICO grammar.

However, to date we have not seen approaches employing a continuous notion of deontics to represent a more fluid understanding of norms as displayed in this work. We believe that our approach is not only useful to represent norm emergence, but also to model long-term adaptation of social institutions, such as transitions between conventions, norms and rules. In this context, note that with its emphasis on different institution types, the approach explored here assumes a higher level perspective on institutions in general, instead of concentrating on specific institution types.

The current work has limitations that will be addressed in future work. Aspects directly related to the dynamic deontics concept include the allocation of terms along a deontic scale, both including a more grounded choice of deontic terms (e.g. 'should') and reviewing the assumption of symmetry along the deontic scale. A further aspect is a more comprehensive consideration of stability/viscosity characteristics, which is currently only applied at the extremes of the deontic scale.

References

1. H. Aldewereld, V. Dignum, and W. Vasconcelos. We ought to; They do; Blame the management! – A conceptualisation of group norms. In *15th International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems*, 2013.
2. G. Andrighetto, D. Villatoro, and R. Conte. Norm internalization in artificial societies. *AI Commun.*, 23(4):325–339, Dec. 2010.
3. A. Bandura, D. Ross, and S. Ross. Transmission of aggressions through imitation of aggressive models. *Journal of Abnormal and Social Psychology*, 63:575–582, 1961.
4. A. P. Cortell and J. W. Davis Jr. Understanding the Domestic Impact of International Norms: A Research Agenda. *International Studies Review*, 2(1):65–87, 2000.
5. S. E. Crawford and E. Ostrom. A Grammar of Institutions. In *Understanding Institutional Diversity*, chapter 5, pages 137–174. Princeton University Press, Princeton, 2005.
6. C. Frantz, M. K. Purvis, M. Nowostawski, and B. T. R. Savarimuthu. nADICO: A Nested Grammar of Institutions. In *PRIMA 2013: Principles and Practice of Multi-Agent Systems*, volume 8291 of *LNAI*, pages 429–436, 2013.
7. A. Ghorbani, H. Aldewereld, V. Dignum, and P. Noriega. Shared Strategies in Artificial Agent Societies. In H. Aldewereld and J. Sichman, editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems VIII*, volume 7756 of *LNCS*, pages 71–86. Springer, Berlin, Heidelberg, 2013.
8. A. Ghorbani, P. Bots, V. Dignum, and G. Dijkema. MAIA: a Framework for Developing Agent-Based Social Simulations. *Journal of Artificial Societies and Social Simulation*, 16(2), 2013.
9. A. Ghose and B. T. R. Savarimuthu. Norms as Objectives: Revisiting Compliance Management in Multi-agent Systems. In H. Aldewereld and J. Sichman, editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems VIII*, volume 7756 of *LNCS*, pages 105–122. Springer, Berlin, Heidelberg, 2013.

10. D. Grossi, H. Aldewereld, and F. Dignum. Ubi Lex, Ibi Poena: Designing Norm Enforcement in E-Institutions. In *Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems II*, volume 4386 of *LNCS*, pages 107–120. Springer, 2006.
11. J. Henrich. Does culture matter in economic behaviour? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon. *American Economic Review*, 90(4):973–979, 2000.
12. A. P. Kinzig, P. R. Ehrlich, L. J. Alston, K. Arrow, S. Barrett, T. G. Buchman, G. C. Daily, B. Levin, S. Levin, M. Oppenheimer, E. Ostrom, and D. Saari. Social Norms and Global Environmental Challenges: The Complex Interaction of Behaviors, Values, and Policy. *Bio-Science*, 63(3):164–175, March 2013.
13. S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan. MASON: A Multiagent Simulation Environment. *Simulation*, 81(7):517–527, July 2005.
14. S. Mahmoud, N. Griffiths, J. Keppens, and M. Luck. Efficient Norm Emergence through Experiential Dynamic Punishment. In *ECAI'12*, pages 576–581, 2012.
15. C. Menger. *Problems in Economics and Sociology*. University of Illinois Press, 1963.
16. J.-J. C. Meyer. A Different Approach to Deontic Logic: Deontic Logic Viewed as a Variant of Dynamic Logic. *Notre Dame Journal of Formal Logic*, 29:109–136, 1988.
17. D. C. North. *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press, Cambridge, 1990.
18. E. Ostrom. *Understanding Institutional Diversity*. Princeton University Press, 2005.
19. S. Panagiotidi, S. Alvarez, and J. Vázquez-Salceda. Towards the Norm-Aware Agent: Bridging the gap between deontic specifications and practical mechanisms for Norm Monitoring and Norm-Aware Planning. In *15th International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems*, 2013.
20. B. T. R. Savarimuthu, R. Arulanandam, and M. Purvis. Aspects of Active Norm Learning and the Effect of Lying on Norm Emergence in Agent Societies. In D. Kinny, J. Y. Jen Hsu, G. Governatori, and A. K. Ghose, editors, *PRIMA*, volume 7047 of *Lecture Notes in Computer Science*, pages 36–50. Springer, 2011.
21. T. Savarimuthu, S. Cranefield, M. A. Purvis, and M. K. Purvis. Obligation Norm Identification in Agent Societies. *Journal of Artificial Societies and Social Simulation*, 13(4):3, 2010.
22. A. Schlüter and I. Theesfeld. The Grammar of Institutions: The challenge of distinguishing between strategies, norms, and rules. *Rationality and Society*, 22:445–475, November 2010.
23. W. R. Scott. Approaching Adulthood: The Maturing of Institutional Theory. *Theory and Society*, 37:427–442, 2008.
24. J. R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, London, 1969.
25. S. B. Shimanoff. *Communication Rules: Theory and Research*. Sage Publications, Beverly Hills, 1980.
26. S. Siddiki, C. M. Weible, X. Basurto, and J. Calanni. Dissecting Policy Designs: An Application of the Institutional Grammar Tool. *The Policy Studies Journal*, 39:79–103, 2011.
27. A. Smajgl, L. Izquierdo, and M. G. A. Huigen. Rules, Knowledge and Complexity: How Agents Shape their Institutional Environment. *Journal of Modelling and Simulation of Systems*, 1(2):98107, 2010.
28. E. Ullmann-Margalit. *The Emergence of Norms*. Clarendon Library of Logic and Philosophy. Clarendon Press, Oxford, 1977.
29. D. Villatoro, G. Andrighetto, J. Sabater-Mir, and R. Conte. Dynamic sanctioning for robust and cost-efficient norm compliance. In *IJCAI'11*, pages 414–419. AAAI Press, 2011.
30. G. H. von Wright. *An Essay in Modal Logic*. North Holland Publishing Company, 1951.
31. G. H. von Wright. *Norm and Action: A Logical Enquiry*. Routledge & Kegan Paul, 1963.
32. C. Watkins. *Learning from Delayed Rewards*. PhD thesis, Cambridge University, 1989.