A Model of Normative Power

Nir Oren, Michael Luck, Simon Miles
Dept. of Computer Science
King's College London
London, UK
nir.oren, michael.luck, simon.miles@kcl.ac.uk

ABSTRACT

A *power* describes the ability of an agent to act in some way. While this notion of power is critical in the context of organisational dynamics, and has been studied by others in this light, it must be constrained so as to be useful in any practical application. In particular, we are concerned with how power may be used by agents to govern the imposition and management of norms, and how agents may dynamically assign norms to other agents within a multi-agent system. We approach the problem by defining a syntax and semantics for powers governing the creation, deletion, or modification of norms within a system, which we refer to as *normative powers*. We then extend this basic model to accommodate more general powers that can modify other powers within the system, and describe how agents playing certain roles are able to apply powers, changing the system's norms, and also the powers themselves. We examine how the powers found within a system may change as the status of norms change, and show how standard norm modification operations — such as the derogation, annulment and modification of norms — may be represented within our system.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*multi-agent systems*

General Terms

Theory

Keywords

Norms, Power

1. INTRODUCTION

Norm aware agents make use of norms, such as obligations, permissions, and prohibitions, to represent and reason about socially imposed goals and capabilities. Such agents are able to decide whether to act in a manner consistent with norms, or whether to ignore them. Norms enable the behaviour of agents (and a system or society as a whole) to be guided or constrained without intruding upon agent autonomy. The ability to modify norms allows a normative system to change as the environment evolves, enabling effective system operation in the face of unexpected situations (by setting up new rules of behaviour); indeed, it has been recognised that self modification is an integral capability for a normative system [3], and increases its robustness and usefulness [1]. It is this ability to modify a norm (norm creation, deletion or change) through agent actions, that we refer to as a *normative power* when the resultant norm is recognised by other agents.

While a number of models have been proposed which support norm modification to a lesser or greater extent (see for example [2, 5]), most of these models ignore the fact that an agent modifying a norm does not have full control of the normative system; that is, it is only capable of modifying a subset of the system's norms, usually in some specific way. For example, a professor within a university has the power to impose some course related obligation or permission on their students, but may not create recognised norms dealing with their university fees. However, it is not clear how the changes brought about by an agent may be represented without enumerating all possible norm modifications in advance. In order to capture this intuition in existing models, one would have to specify all possible norm modifications in advance. This is clearly not practical in open domains, where new situations may arise, and new norms may appear as the system evolves. Even in closed domains, specifying all possible norm modifications may be infeasible due to the number of modifications that may need to be specified. In this paper, therefore, we focus on a mechanism for system self modification built around the action of constituent agents.

Our contributions in this paper are threefold. First, we provide a framework for norms in which the application of normative power can be modelled. That is, given an initial set of norms, we identify those norms that exist after an agent has taken some norm modification related action. Second, we describe a method for restricting the types of norms an agent may modify, and the ways in which such modification can take place (thus providing control over how the normative system may be changed). Finally, our analysis can be applied not only to norms, but also to normative powers, and to powers over such powers ad-infinitum, allowing us to specify the powers that an agent may modify in the system. We term such powers *general powers*, and show how our model can be used to describe the effects of an application of a general power, and restrict the general powers an agent may make use of. We show the applicability of our work by demonstrating how it may be applied to model common situations such as when a norm must be *derogated*, and examine how norms and powers may be dynamically modified within an open environment.

2. NORMS

Before considering norms in detail, we introduce the motivating example of an academic environment consisting of a number of professors, post-doctoral researchers, and students, each working

Cite as: A Model of Normative Power, N. Oren, M. Luck, S. Miles, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. XXX-XXX.

on various projects. In such a situation, professors may be obliged to teach certain courses, while post-docs and students may need to write papers according to the demands of the professor running their project. Norms to describe this type of institution include those that require professors to give lectures on certain courses, students to write papers, and so on.

Professors may have certain powers. For example, they may be able, but are not required, to oblige a post-doc to replace them when giving a lecture, or may oblige a student or post-doc to write a paper, possibly for a specific conference.

Our model of power is closely aligned with the normative model originally proposed in [10], designed to track complex changes to the status of norms over time, and is particularly applicable to contract representation and reasoning. Below, we provide a brief outline of this normative model.

At the heart of the model lies the concept of a norm, which identifies an obligation or permission that comes into force in a particular situation, and which ceases to affect an agent given some other situation. The norm's components are represented in some logical predicate language $\mathcal{L}$. The choice of a specific predicate language is left open as the model only requires abstract concepts that can be specialised according to the particular system to which the norms apply, as well as the associated languages used for representation and reasoning. In this paper, we assume that $\mathcal{L}$ is first order logic.

Norms may be *abstract* or *instantiated*. Abstract norms identify the classes of situations in which obligations or permissions may come into force or cease to be, while instantiated norms represent the particular obligations or permissions applicable to sets of agents when the situations identified by the abstract norms occur. Abstract norms consists of five elements: a norm type (NTy) stating whether the norm is an obligation or permission[1]; an activation condition NA, identifying the conditions under which it is activated (i.e. instantiated); the class of agents (the norm's *targets*, NTa) to which the norm applies when activated; an expiration condition NE, identifying the conditions under which an instantiated norm expires (i.e. no longer has normative force); and the states of affairs (known as the *norm conditions*, and labelled NC below) that are obligatory or permitted (depending on the norm's type).

DEFINITION 1. ***(Norm)*** *A* norm $\mathcal{N}$ *is a tuple of the form*

$$\langle NTy, NA, NC, NE, NTa \rangle$$

where: $NTy \in \{obligation, permission\}$*, and* $NA, NC, NE,$ *and* NTa *are all* wff *in* $\mathcal{L}$ □

Thus, we could have an abstract norm obliging a professor P, who runs a course C to give a lecture at the appropriate time on the subject, and that the lecture must take two units of time. More specifically, if P gives course C, and the time the course starts is Y, and it is now time Y, then, until the current time is $Y+2$, the lecturer is obliged to be giving a lecture L, with the subject of the lecture being the course. [2]:

$$\begin{aligned}\textbf{LectureObligation}: \quad & \langle obligation, \\ gives Course(P,C) \wedge timeOfCourse(C,Y) & \wedge \\ currentTime(T_1) \wedge T_1 = Y&, \\ givingLecture(P,L) & \wedge \\ subjectOfLecture(L,C)&, \\ currentTime(T_2) \wedge T_2 > Y+2&, \\ professor(P) & \rangle\end{aligned}$$

[1]Prohibitions are obligations with negated normative conditions.

[2]We make use of Prolog notation within our logical formulae. Variables begin with a capital letter, and constants with lowercase.

In this case, the abstract norm to give a lecture holds whenever the current time equals the course's time (i.e. the norm's NA parameter holds). When this occurs, the norm affects the agent (in this case obliging it to give the lecture), and the norm is thus instantiated. More generally, if, at any point, an abstract norm's NA holds, an instantiated version of the norm is created (subject to constraints discussed in [10]). This instantiation involves creating a copy of the abstract norm in which the norm's variables are bound to the values that caused the activation condition to evaluate to true. When instantiated, the individuals specified by NTa are identified. These individuals are then either obliged or permitted (as determined by NTy) to bring about the normative goal specified by NC, until the conditions specified by NE hold.

Thus, a norm is *instantiated* with respect to some background logical theory $\mathbf{\Gamma}$ describing the environment, if NA can be inferred from $\mathbf{\Gamma}$, and NTa can be inferred from $\mathbf{\Gamma}$ as a ground formula (i.e, all variables in NTa are instantiated). Given the above norm, if a professor *smith* gives a course on *physics*, and the time of the course is 11, (*givesCourse(smith,physics)* and *timeOfCourse(physics,11)* are true, as well as *professor(smith)*), then we obtain the following instantiated norm:

$$\begin{aligned}\textbf{LectureObligation}': \quad & \langle obligation, \\ givesCourse(smith, physics) \wedge timeOfCourse(physics, 11) & \wedge \\ currentTime(11) \wedge 11 = 11&, \\ givingLecture(smith, L) & \wedge \\ subjectOfLecture(L, physics)&, \\ currentTime(T_2) \wedge T_2 > 13&, \\ professor(smith) & \rangle\end{aligned}$$

It should be noted that while the entire activation condition is ground within an instantiated norm, the normative and expiration conditions may contain variables. These conditions evaluate to false until some grounding is possible, from formulae obtained from the domain, which cause the condition to evaluate to true.

Building on this base, [10] describes how certain predicates related to the *status* of a norm may be evaluated. In particular, a norm's status relates to concepts such as being *instantiated*, *violated* and *expired*. Its status may change over time, and the model captures this notion by proposing a simple temporal structure from which the status of norms may be derived at any time. Using this structure, predicates may be created referring to a norm's status. For example, the predicate $violated(\mathcal{N}, t)$ evaluates to true if norm $\mathcal{N}$ is violated at time t. These predicates may be used like any others, and can thus be thought of as forming part of $\mathbf{\Gamma}$. We assume that all predicates within our system are temporal in nature, and evaluate to true, or false, at different points in time. We thus write $\mathbf{\Gamma}^t$ to represent the subset of $\mathbf{\Gamma}$ which contains all predicates true at time t, and label the instant following it $\mathbf{\Gamma}^{t+1}$.

3. NORMATIVE POWER

The normative model described in the previous section allows an agent to determine what it must, may, or may not do. It also allows an agent to determine whether some norm was or was not complied with. However, it does not make any attempt to model how a set of norms may be assigned to an agent, modified, or deleted, and whether such operations should be complied with by the agent.

Thus, for example, the normative model cannot represent the ability of a professor to create a new norm, requiring a student to write a paper for some conference. Clearly, if such a norm was to be created, it should be complied with by the student. However, a norm representing another demand by the professor, for example obliging the student to wash the professor's car, would not be recognised as a valid obligation by the student, who would not incur any penalties from the institution if it chose to violate this obligation.

In this section, therefore, we consider how we may capture the notions underlying the modification of norms in a system through exercising power over those norms; we introduce a new model of normative power. We begin with an informal analysis of power, after which we detail a simple model of power, able to capture the notion that an agent has the power to affect a single situation. We then refine this model by describing *norm templates*, which identify sets of related norms that may be affected by a single power. Note that the examples used in the remainder of this paper are somewhat contrived for reasons of exposition and berevity, but we believe clarify the points we are trying to illustrate. In particular, we avoid the notion of deadlines but have illustrated their use in Section 3.

3.1 The Notion of Power

An agent able to create, delete or modify some norms within an institution is said to have *normative power* over those norms.

In their seminal work, Sergot and Jones [8] make mention of three distinct aspects of power, namely:

1. a legal power, which is *recognition by the institution* that some action will have some (institutional) effect;
2. a physical power, which is the *ability of an agent* to undertake some action that will have the power's effect; and
3. a *permission* to undertake the action.

In this paper, the notion of normative power adopted is a combination of the first two of these aspects. That is, an agent with normative power over some norm has the capability to alter the norm in some way that is recognised by the appropriate institution (i.e. the power is legal). The third aspect here is not core to our model of power, but we do consider it further in Section 6.

3.2 A Model of Power

Suppose we have the situation in which a professor, John, wants their student, Isaac, to write a paper for AAMAS. Given the normative model described in Section 3, we can express this norm (which we label *WP*, for *writing a paper*) as follows.

$$\begin{aligned}\mathbf{WP}: \quad & \langle obligation, \\ & \top, \\ & writingPaper(A) \wedge target(A, aamas), \\ & written(A), \\ & student(isaac)\rangle\end{aligned}$$

Here, $\top$ indicates that the norm is always applicable. In other words, it will immediately be instantiated.

In order to capture what is required to represent John's power to impose the *WP* norm (in this case, by creating it), we need to identify John as the agent exercising the power, and *WP* as the norm being imposed. More generally, we often also want to specify *when* this power can be exercised, and the particular modifications to norms that are covered by the power. The former of these can be represented using a condition, as in the case of the activation and expiration conditions of norms. The latter is more interesting in that it should permit the addition of new norms (as in our example), the deletion of norms, and the modification of existing norms. To represent this, we consider the norms before and after the modification.

Given our representational requirements, we need to be able to identify the following in a model of power:

1. The agents that are able to exercise the power.
2. The situations in which the power may be applied.
3. The norms or powers that must already exist in the system, and which are modified or removed by the exercise of the power.
4. The norms or powers that *will* exist in the system after the application of the power, *in place of* the norms or powers identified by the previous point.

This provides us with the basis for an initial definition of normative power, as follows.

DEFINITION 2. ***(Naive Normative Power)*** *A normative power is a tuple*

$$\langle Mandators, Context, Pre, Post\rangle$$

where $Mandators$ is a set of predicates identifying agents; $Context$ is a logical sentence identifying the situations in which the power may be applied; Pre and $Post$ are sets of norms removed and inserted within the system by the application of the power. □

Making use of this definition gives rise to a data structure, for our example of a professor (John) having the power to make their student (Isaac) write a paper for AAMAS, as follows.

$$\begin{aligned}&\langle professor(john), \\ &\top, \\ &\{\}, \\ &\{WP\}\rangle\end{aligned}$$

Here, the first element identifies the agent(s) and roles able to exercise the power; the second identifies the situations in which the norm may be applied (with $\top$ stating that the power may be applied in all situations); the third parameter lists the set of norms that are deleted from the system during power application; and the fourth lists those norms that appear in the system after the power is applied. Note that we have written WP within the power for conciseness; formally, the entire WP norm should appear in place of WP within the power.

Based on this pattern, norm creation then amounts to a similar structure with an empty set of deleted norms in the third parameter of the power, and a set of newly created norms in its fourth parameter. Conversely, norm deletion amounts to listing norms to be deleted in the third parameter, and leaving the fourth empty. Finally, norm modification can take place by putting norms in both the third and fourth parameters.

However, this approach of explicitly listing each of an agent's normative powers, for each relevant norm, is not practical. Returning to our example, suppose that John the professor also requires Isaac the student to write a paper for IJCAI. Here, we need an additional power, similar to the first, with a norm WP' instead of WP, using the predicate $target(X, ijcai)$ to target the new conference. Clearly, particularly in an open environment where new conferences may appear, an exhaustive list of powers targeting various conferences is undesirable (or even impossible to generate). Instead, we would like to be able to specify a power that allows a professor to oblige their students to write a paper for some conference, but without specifying which conference. Intuitively, if someone has a normative power then they should be able to modify a *set* of relevant norms in some way, rather than just focusing on an individual norm.

$\mathbf{N1}: \quad \langle obligation,$ $wantsPaper(T, P) \wedge supervisor(S, T),$ $writtenPaper(S, P),$ $writtenPaper(S, P),$ $student(S)\rangle$
$\mathbf{N2}: \quad \langle obligation,$ $wantsPaper(T, P) \wedge supervisor(S, T) \wedge target(P, C),$ $writtenPaper(S, P) \wedge submitted(P, C),$ $writtenPaper(S, P) \wedge submitted(P, C),$ $student(S)\rangle$

Table 1: A general (N1) and more specific (N2) norm.

3.3 Norm Templates

One way to capture the intuition of modifying a set of norms rather than just an individual norm is to specify a *norm template*, with the aim of identifying the families of norms affected by the power. In this model, if an individual norm fits the norm template of some power P, then that normative power may affect the norm in some way.

A norm template is thus a general form of norm suitable for identifying families of norms; it can be defined to be the same as a norm, but with the possibility of generalising the norm type to include both obligations and permissions, for example (thus allowing for it to capture both permissions and obligations related to some set of norm targets and activation, normative and expiration conditions).

DEFINITION 3. ***(Norm Template)*** *A norm template is a norm with one possible additional element in NTy, namely the truth symbol $\top$.* □

Norm templates may be more or less general but, given the structure of norms and their conditions, clearly determining whether a norm fits a norm template is not trivial. For example, a norm requiring students to *write papers as specified by their supervisor* is more general than one requiring students to *write a particular paper for a particular conference, as specified by their supervisor*. These norms can be represented as $N1$ and $N2$ in Table 1. In a sense, the latter norm may be viewed as a special case of the former norm. Furthermore, if some authority is able to issue the former norm, then they should be able to issue the latter. Our procedure for determining whether a norm fits a norm template is built around this idea of norm specialisation, or *norm subsumption*. Therefore, we identify a procedure for inferring whether one norm (or norm template) *subsumes* some other norm (or norm template).

Given our representation of norms, one norm, N1, is more general than another, N2, if N1's activation condition, normative condition, and expiration condition can each be inferred from (the more specific norm) N2's activation, normative and expiration condition. However, as we discuss below, some refinements to this approach are needed in the case of the normative condition. Before discussing these refinements, we examine the activation and expiration conditions more closely.

3.3.1 Activation and Expiration Conditions

Consider two norms, one with an activation condition of x, and another with an activation condition of the form $x \wedge y$, for some formulae x and y. These activation conditions can be read as stating "if x then ...", and "if x and y then ..." respectively. Clearly, the latter activation condition is more specific than the former, and any power which allows the former norm should allow the latter. In other words, specialisation of the antecedent via conjunction introduction is valid. Similarly, an expiration condition containing $x \wedge y$ means "... until x and y occur", which again, is valid to infer as a specialisation of an expiration condition of the form x. Thus, the activation and expiration conditions of norm $N2$ in Table 1 are clearly specialisations of these conditions in norm $N1$. Note also that given a norm with some activation condition x, we may not infer that some other norm with activation condition y is a specialisation of that norm, as x and y are unrelated. Thus, a norm of the form "if it is the weekend, you must work" is not one that should be inferred as a specialisation (or generalisation) of a norm stating "if it is a weekday, you must work". Given this analysis, it is clear that standard logical inference may be used to determine whether some activation or expiration condition is a specialisation of some other activation or expiration condition.

3.3.2 The Normative Condition

However, this analysis does not hold for the normative condition. In our example, the power enabling a professor to ask a student to write a paper and target it at the AAMAS conference (expressed in the normative condition), is a valid application of the more general power to require a student to write a paper. Using $\vdash$ to represent inference, we may conclude that

$$writingPaper(S, P) \wedge submitTo(P, C) \vdash writingPaper(S, P)$$

However, by using a logic that allows for conjunction elimination, we must also admit the following inference:

$$writingPaper(S, P) \wedge stealFrom(S, U) \vdash writingPaper(S, P)$$

In other words, by using this simple analysis, the application of a power allows for a norm condition to oblige the party affected by the norm to achieve some state of events *and* any other state of events. This is clearly both undesirable and inappropriate.

It appears that, unlike the expiration and activation conditions, there is no way to generalise an arbitrary logical formula's form in a way that is consistent across different norms. However, it is clear that many generalisations of a norm are ontological in nature, in the sense that a norm may refer to some more specific concepts than exist in some other, more general norm. For example, the concept of *writing a paper* may be a subconcept of the more general concept of *doing academic work*. A professor able to require a post-doc to *do academic work* should thus also be able to require the post-doc to *write a paper*. Without dwelling on the vast literature on ontologies, we capture the notion of ontological generalisation through a simple predicate hierarchy where $p1(\ldots) \leq p2(\ldots)$ denotes that $p1$ is more than, or as specific as $p2$, and use this in defining the notion of norm subsumption, in which one norm is identified as more general than another. We may extend this notion to logical sentences; we write $P \leq Q$ iff for every predicate p in P, substituting it with a predicate q such that $p \leq q$, we obtain Q.

In order to capture ontological specialisation in our model, we define a new inference operator $\Rightarrow$ to include both logical inference, as well as the ontological dimension of subsumption. Thus, $A \Rightarrow_O B$ iff $A' \vdash B$ where $A \leq A'$ with respect to some ontology O. In what follows, we implicitly assume the existence of an ontology, and thus write $A \Rightarrow B$.

Finally, we may also want to specify that an agent may have unlimited scope regarding some norm parameter, for example *any* activation condition for the norm. We make use of the $\top$ symbol as a wildcard to capture this notion, and reflect this in our definition of norm subsumption.

3.3.3 Norm Subsumption

We are now in a position to define norm subsumption. One norm template $n1$ subsumes another norm or norm template $n2$ (i.e. $n1$ is more general than $n2$) if they share a norm type, or the former's type is $\top$; $n1$'s activation and expiration conditions may be inferred (with respect to some ontology) from $n2$; and $n1$'s normative condition and norm target are ontologically more general than $n2$'s. Alternatively, if any parameter in $n1$ is $\top$, that parameter in $n2$ is automatically assumed to satisfy the condition for subsumption.

DEFINITION 4. ***(Norm Subsumption)*** *Given a partially ordered hierarchy of predicates, and two norm templates $n1, n2$ of the form*

$$\langle NTy_i, NA_i, NC_i, NE_i, NTa_i \rangle$$

where $i \in \{1, 2\}$ respectively, we say that $n1$ subsumes $n2$ (i.e. $n1$ is more general than $n2$), written $n1 \gg_n n2$ iff

1. $NTy_1 = \top$ *or* $NTy_1 = NTy_2$
2. $NA_1 = \top$ *or* $NA_2 \Rightarrow NA_1$
3. $NE_1 = \top$ *or* $NE_2 \Rightarrow NE_1$
4. $NC_1 = \top$ *or* $NC_2 \leq NC_1$ *or* NC_2 *is a ground version of* NC_1
5. $NTa_1 = \top$ *or* $NTa_1 \leq NTa_2$ □

Thus, for example, a norm with an activation condition a is subsumed by a norm with an activation condition $a \vee b$, while a norm with an activation condition $a \wedge b$ subsumes a norm with an activation condition a. Now, we may revise our earlier definition of naive normative power, using norm templates rather than norms.

DEFINITION 5. ***(Normative Power)*** *A normative power is a tuple*

$$\langle Mandators, Context, Pre, Post \rangle$$

where $Mandators$ is a set of predicates identifying the agents able to exercise the power; $Context$ is a logical sentence identifying the situations in which the power may be applied; and Pre and $Post$ are sets of norm templates *identifying the norms removed and inserted within the system by the application of the power.* □

4. APPLYING POWERS

While we have defined the structure of a normative power, we have not yet considered the effects of the application of the power on the norms within an environment.

4.1 Institutional Environment

So far, we have considered norms and powers as disembodied concepts, but they must exist somewhere. Conceptually, they may be considered as mental components within individual agents which themselves make up an institution. We abstract away from the individual agents by representing norms and powers as belonging to the institution, and assume that this institution can change over time. For convenience, we assume a discrete time model, and thus, at any time point t, there exists some set of norms and powers within the institution. The application of a power at time t then alters the norms that exist at time $t + 1$. Changes in the environment and agent actions may also alter the norms over time, due to the instantiation of new norms from abstract norms, and the expiration of other instantiated norms due to agent action. We refer to the set of norms and powers as the *institutional environment*.

The institutional environment may be viewed as a trace, and is used to track the powers found within the institution at any point in time, as well as the abstract norms imposed on the members of the institution. The institutional environment thus stores the *abstract norm set* $\mathcal{ANS}$ from which norms are instantiated within the normative environment. We are now in a position to define the institutional environment.

DEFINITION 6. ***Institutional Environment*** *The institutional environment IE is a set of triples*

$$\langle T, \mathcal{ANS}, Powers \rangle$$

where $T \in \mathbb{Z}$ identifies a time point, $\mathcal{ANS}$ is a set of abstract norms, and $Powers$ is a set of powers.

A norm n is said to be within an institutional environment IE at time T iff $\langle T, \mathcal{ANS}, Powers \rangle \in IE$ and $n \in \mathcal{ANS}$.

A power p is said to be within an institutional environment IE at time T iff $\langle T, \mathcal{ANS}, Powers \rangle \in IE$ and $p \in Powers$.

The roles able to exercise a power, within the context of the institutional environment, are captured within the power's mandator's. A mandator of the form *professor(X)* means that any agent in the *professor* role is able to exercise the power.

4.2 Core Power Operations

Given this, we can define three basic operations using normative powers: the generation, deletion and modification of norms. In what follows, we examine how the application of a power to perform these operations affects the institutional environment.

Generating Norms Given an institutional environment at time t containing a power of the form

$$\langle A, C, \{\}, N \rangle$$

where A is a set of agents, C is a logical formula that is true at time t, and N is a set of norm templates where, for at least one $n \in N$, there is no norm in the institutional environment that is subsumed by n, the application of the power results in an institutional environment at time $t + 1$ where, for each $n \in N$, there is a norm that is subsumed by n.

Deleting Norms Given an institutional environment at time t containing a power of the form

$$\langle A, C, N, \{\} \rangle$$

where A identifies a set of agents, C is a logical formula that is true at time t, and N is a set of norm templates such that for all $n \in N$ there is a norm in the institutional environment that is subsumed by n, application of the power results in an institutional environment at time $t+1$ in which, for each $n \in N$, there is no norm that is subsumed by n. (This is clearly a very strong operation, deleting *all* norms subsumed by n; deleting just one norm may be desirable, but it is not clear how this might be determined, and we leave consideration of this to future work.)

Modifying Norms The modification of a norm consists of deleting some norms in the system, and creating other norms in their place. Thus, given an institutional environment at time t, containing a power of the form

$$\langle A, C, N_a, N_p \rangle$$

$\mathbf{delayPower}: \langle conferenceOrganizer(CO), \top, \{submitObligation\}, \{submitObligation, delayPermission\}\rangle$
$\mathbf{submitObligation}: \langle obligation, dateToday(21August), submittingPaper(P), paperSubmitted(P), agent(A)\rangle$
$\mathbf{delayPermission}: \langle permission, dateToday(21August), \neg submittingPaper(P), paperSubmitted(P) \vee dateToday(28August), agent(A)\rangle$

Table 2: An illustration of derogation.

then application of the power results in an institutional environment at time $t+1$ in which first power $\langle A, C, N_a, \{\}\rangle$ is applied and then power $\langle A, C, \{\}, N_p\rangle$ is applied.

4.3 Complex Power Operations

Norms may be modified in several different ways. Building on the core power operations specified above, we can derive further operations. For example, *derogation* is an operation sometimes considered in the context of norms when they are not fully revoked. In our model, we can view this as the creation of a permission to override some obligation. For example, norm $submitObligation$ in Table 2 states that an agent may be obliged to submit a paper on the 21st of August. Power *delayPower* states that a conference organiser may create a permission allowing submission a week later (represented by the permission $delayPermission$ within the Table 2).

If, at time t_0 the set of norms contains $submitObligation$, the application of the power (which may be applied by the conference organiser at any time, due to the context condition $\top$) will result in the set of norms containing both $submitObligation$ and $delayPermission$ at time t_1. Thus, while the obligation to submit a paper on the 21st of August still exists, any agent not submitting a paper on that date will not be violating the obligation due to the existence of the permission (where permissions are treated as exceptions to obligations). It should be noted that the permission's activation condition means that the power must be used before, or on the deadline date, otherwise the permission will not come into force.

In general, some obligation O with normative condition NC_O may be derogated by creating a permission P with normative condition $\neg NC_O$. As seen above, the pattern for such a power is of the form $\langle M, C, \{O\}, \{O, P\}\rangle$. While the original obligation is deleted, it is reinserted due to the power's postconditions. Thus, as expected the power may not be applied if the obligation does not currently exist. The scope of the derogation may be controlled through the permission's activation and expiration conditions.

5. GENERAL POWER

In a manner similar to the way in which agents may create, delete, or modify a norm, agents may also be able to create, delete or modify powers (which may themselves be powers over norms or over other powers). We refer to such a power as a *general power*. The act of performing the operations described by a power is called *exercising* or *applying* the power. For example, a professor may give a post-doc the power to ask a PhD student to write a paper.

We treat general powers as an extension to normative powers, capturing both the notion of general and normative power (and thus replacing the latter). We define a general power as a power over a power or over a norm.

DEFINITION 7. ***(General Power)*** *A general power gp, is a tuple $\langle M, C, Pr, Po\rangle$, containing M a set of predicates identifying agents; C a logical sentence identifying the situations in which power may be applied; and Pr and Po, which are pairs of the form $\langle GPT, NT\rangle$, where GPT is a set of general powers, and NT is a set of norm templates.* □

The pairs contained within Pr and Po include sets of general powers. Each element of these sets may in turn hold further general powers, and if left unconstrained, this may continue ad infinitum. We therefore require that only a finite number of general powers exist within the system, and also that a general power cannot contain a reference to itself within its Pr or Po elements, no matter how far removed. That is, if a general power GP_1 contains another general power GP_2 within the set of general powers found in its Pr or Po elements, and this GP_2 contains a general power GP_3 with its Pr or Po elements, and so on until some general power GP_n (which contains no further general powers), then all general powers $GP_1, \ldots, GP_n$ must be unique.

Subsumption over a general power occurs when both the powers and norm templates within Pr and Po are subsumed. Since the requirements described above prohibit the possibility of loops of general power, we may define subsumption recursively as follows.

DEFINITION 8. ***(Subsumption over General Powers)*** *Given two general powers*

$$GP_1 = \langle M_1, C_1, Pr_1, Po_1\rangle$$
$$GP_2 = \langle M_2, C_2, Pr_2, Po_2\rangle$$

We say that GP_1 subsumes GP_2, written $GP_1 \gg_g GP_2$ iff

1. $M_2 \subseteq M_1$
2. $C_2 \Rightarrow C_1$
3. *If $Pr_1 = \langle GPT_1, NT_1\rangle$ and $Pr_2 = \langle GPT_2, NT_2\rangle$, then $\forall t_1 \in GPT_1, \exists t_2 \in GPT_2$ such that $t_1 \gg_g t_2$. $\forall t_1 \in NT_1, \exists t_2 \in NT_2$ such that $t_1 \gg_n t_2$.*
4. *If $Po_1 = \langle GPT_1, NT_1\rangle$ and $Po_2 = \langle GPT_2, NT_2\rangle$, then $\forall t_1 \in GPT_1, \exists t_2 \in GPT_2$ such that $t_1 \gg_g t_2$. $\forall t_1 \in NT_1, \exists t_2 \in NT_2$ such that $t_1 \gg_n t_2$.* □

5.1 Applying General Powers

Using a similar approach to that used in Section 4, we can show the effects of the application of a general power on the structure of norms within the environment.

Given a general power $\langle A, C, Pr, Po\rangle$ such that the institutional environment contains the general power at time t and C evaluates to true, the power is applied iff the following conditions hold.

- At time t, given $Pr = \langle GPT, NT \rangle$, for any elements $gpt \in GPT$, there is a general power within the institutional environment that is subsumed by gpt. For any $nt \in NT$, there is some norm within the institutional environment that is subsumed by NT.

  Given $Po = \langle GPT', NT' \rangle$, for any element $gpt' \in GPT'$, there is no general power within the institutional environment that is subsumed by gpt' unless it is subsumed by an element of GPT. For any $nt' \in NT'$, there is no norm within the institutional environment that is subsumed by nt' unless it is subsumed by an element of NT.

- At time $t+1$, given $Pr = \langle GPT, NT \rangle, Po = \langle GPT', NT'$, for any element $gpt \in GPT$, there is no general power within the institutional environment that is subsumed by gpt unless it is subsumed by an element of GPT'. For any $nt \in NT$, there is no norm within the institutional environment that is subsumed by nt unless it is subsumed by an element of NT'.

  For any element $gpt' \in GPT'$, there is a general power within the institutional environment that is subsumed by gpt'. For any norm $nt' \in NT'$, there is a norm within the institutional environment that is subsumed by nt'.

Following [5], we can represent the application of a power by a predicate within the environment, allowing it to be used within norms. Thus, if

$$\mathbf{\Gamma}^t \vdash apply(A, P)$$

then power P was applied by agent A at time t.

5.2 Powers over Powers

As for normative powers, we may identify a number of stereotypical operations over general powers, such as the creation and deletion (*annulment*) of a power. In addition, a particular *instantiation* of a power is delegation, wherein a power is transferred to some other agent, which may possibly be able to further transfer this power. For example, if Bob is obliged to pay Alice five dollars, and Alice would like to oblige Bob to give the money to Charlie instead, she must, after changing the obligation, transfer the ability to further transfer the obligation to Charlie as well (as Charlie could, for example, ask Bob to pay Doris the money instead). At this point, two possible interpretations of the situation are possible. First, it could be the case that once the obligation is transferred, Alice may no longer affect it. She could thus not later ask for the money to be paid to herself rather than Charlie again. Thus, the transfer of power means that Alice loses all power over the obligation. The second interpretation allows her to transfer the power back to herself, meaning that Charlie does not have as much power over the obligation as Alice, who can continue affecting it, even after the application of the power.

5.2.1 Infinite Power Sequences

In the first case, it would appear that Alice must not only transfer the obligation, and the power over the obligation, but also the power over the power over the obligation, and so on ad infinitum. Since we assume only a finite number of powers, such *infinite power sequences* appear to be a weakness. However, since a norm (or power) must exist within the system in order for a power affecting it to be applied, such infinite power sequences do not actually arise in practise. For example, we may write the norm above as

$$N_{alice} = \langle obligation, true, pay(alice, 5), paid(alice, 5), agent(bob) \rangle$$

where the subscript identifies the target of the norm. By associating a power of the form

$$P_X = \langle agent(X), C, \langle \{\}, N_X \rangle, \langle \{\}, N_Y \rangle \rangle$$

with every agent in the system (where X and Y represent wildcards able to refer to every agent name, and thus N_{alice} represents the norm defined above), the ability to transfer a norm without resorting to infinite power sequences is provided. Using the example of Alice and Charlie, Alice would have to apply the power P_{alice}, substituting Charlie for Y.

In order to represent the second situation, a modification of the P_X power is required: an agent may only transfer a norm if they had transferred it before.

$$P'_X = \langle agent(X), apply(X, P'_X), \langle \{\}, N_Y \rangle, \langle \{\}, N_Z \rangle \rangle$$

Clearly, some bootstrapping mechanism is needed before this power can be applied. This can take the form of another power, allowing the original target of a norm to transfer the power. However, determining a norm's target requires examining a norm's historic state, and is outside the scope of this paper.

5.2.2 Permission to Apply Powers

As mentioned in Section 3.2, Jones and Sergot [8] identified three distinct aspects of power. It is clear that our notion of normative and general power capture the first two of these aspects, namely the legal aspect, which states that recognition of the application of the power will have some institutional effect, and the physical aspect, capturing the agent's ability to make use of the power. The third aspect involves the permission to make use of the power, and is not captured by normative and general powers. It is easy to imagine a situation where an agent could apply a normative power in such a way that it is recognised by the relevant institution, but do so in a way that it is not permitted to act. For example, a professor could contravene university regulations and oblige a university to grant a failing student permission to enter the next year of study. In order to capture the third aspect of power, we must return to the normative model, and create a permission for an agent to exercise some power. In order to do so, we use the $apply(A, P)$ predicate within the normative context of a permission (or an obligation, or a prohibition as a negated obligation) to show that an agent may (or must, or may not) make use of the power under some circumstances. Similarly, the application of a power may have normative side effects, and the use of the *apply(...)* predicate within a norm's activation condition may be used to impose these side effects on an agent.

6. DISCUSSION

In this paper, we presented a model of power which is able to directly affect norms and other powers. An alternative interpretation of our model could be imagined, where the application of a power makes some certain institutional fact true [8], and the presence of these institutional facts affects which norms are in force, and which additional institutional facts may then be asserted. However, as discussed in Section 3.3, identifying whether a power can or cannot be applied requires non-standard inference to be performed, requiring a new logic to be defined.

Other researchers have examined issues relating to norm modification and normative power. In [5], Gelati *et al.* describe *declarative power*, as "the capacity of a power holder to create normative positions ... by proclaiming such positions". They provide a logical semantics for declarative power, and link it to standard action logics. However, they point out that their model cannot easily represent limitations on power, or how a system in which declarative power exists may change over time. Similarly, work such as [4] examines only an instant of time, and thus cannot model the evolution of a normative system under the influence of power.

Building on the notion of a proclamation, [6] and [7], Governatori investigates modifications of a normative system using temporal extensions to defeasible logic. This work is based on the idea of a *norm modification function*, which takes in a norm valid at some time, and returns another norm together with the time at which it is valid. The act of norm modification then makes use of this function, replacing the norm passed into the function with the function's output. Governatori identifies four specific norm modifications, namely annulment, partial and total substitution, and derogation, and shows how these may be represented as predicates. In our framework, annulment corresponds to the deletion of a norm, while partial and total substitutions are equivalent to our norm modifications. Derogation can be modelled in our framework through the introduction of a new permission, while Governatori's work focuses on proof procedures, allowing an agent to determine the rules and literals that may be derived given an initial set of rules and facts. Here, proclamations serve as inputs to certain pre-created rules, but Governatori does not discuss the structure, nor examine how limits to an agent's power may be defined.

In [2], Boella *et al.* provide an in depth examination of different types of permissions. They show that using permissions, it is possible to limit the changes that a lower level authority can perform on a normative system. In order to limit the permissions an authority can provide to another agent, they introduce the notion of a *competence*; an agent may only issue a permission to another agent if the permission falls under the issuer's competence. Their norms are represented as I/O logic [9] $(antecedent, consequent)$ pairs, and this notion of competence is defined only on the consequent of a norm, meaning that while it is possible to model some types of restrictions on an agent's powers, restrictions based on a norm's antecedents cannot be represented.

One issue we have not yet examined is how simultaneous modifications of norms or powers should be handled. Since any application of power deletes the original norm or power from the system before creating a new version at the next point in time, simultaneous actions will cause both operations to take place. However, this may be undesirable in some situations (for example if an agent were to delete a power while another were to modify it, some would argue that the modification should fail). In this case, the addition of a priority relation between powers would allow powers to be applied in some given order, preventing these counter-intuitive results.

Finally, it should be noted that the framework of Oren et al. [10] was used due to its underlying similarities with our model of power. However, we believe that the work presented here can be adapted to represent the notion of power in other normative models.

7. CONCLUSIONS AND FUTURE WORK

In this paper we introduced the notions of normative and general powers. Normative powers express the fact that one or more agents have the ability to add, delete, or modify a norm, possibly under some specific situation. General powers represent the ability of an agent to add, delete, or modify a normative or other general power (in some specific situation). We showed how an agent's application of a power alters a set of norms and powers, and showed how an agent's powers may be restricted.

The concept of a power is distinct, but related to, other normative concepts such as obligations and permissions. By introducing normative power into a multi-agent system, agents may dynamically create new norms, allowing for increasingly complex interactions to take place, while still placing bounds on system behaviour. This increased richness results in more flexible norm-aware agents, and allows for the modelling of more complex domains.

One issue that we have not yet examined is the concept of system specification. For example, what constraints on powers are required so as to guarantee that no conflicting norms will ever appear in a system? Similarly, is there a way to identify the simplest set of norms and powers required so as to guarantee that some desirable behaviour is achieved? Such questions are common in multi-agent systems research, and are usually answered by creating a logic to represent the framework. In future work we will investigate how our framework may be extended to answer these questions.

The distinction between normative power and simply giving orders is both subtle and controversial, which we will explore further. Additionally, we intend to extend our power model to allow for a finer grained representation of normative power, focusing on the limitations of our approach in specifying normative conditions. We also intend to extend our normative framework so as to be able to represent concepts such as a norm's creditor directly.

8. ACKNOWLEDGEMENTS

The work reported in this paper was funded by the Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre established by the UK Ministry of Defence

9. REFERENCES

[1] G. Boella, G. Pigozzi, and L. van der Torre. Normative systems in compute science, ten guidelines for normative multiagent systems. In *Dagstuhl Seminar Proceedings 09121*, 2009.

[2] G. Boella and L. van der Torre. Permissions and obligations in hierarchical normative systems. In *Proc. ICAIL '03*, 2003.

[3] C. Castelfranchi. Social power. a point missed in multi-agent DAI, HCI. In Y. Demazeau and J. P. Müller, editors, *Decentralized A. I.*, pages 49–62. Elsevier Science, 1990.

[4] R. Demolombe and V. Louis. Norms, institutional power and roles: Towards a logical framework. In *Proc. of the 16th Int. Symp. on the Foundations of Intelligent Systems*, pages 514–523, 2006.

[5] J. Gelati, G. Governatori, A. Rotolo, and G. Sartor. Normative autonomy and normative co-ordination: Declarative power, representation, and mandate. *Artificial Intelligence and Law*, 12(1-2):53–81, 2004.

[6] G. Governatori, V. Padmanabhan, and A. Rotolo. Rule-based agent in temporalised defeasible logic. In *Proc. PRICAI 2006*, volume 4099 of *Lecture Notes in Artificial Intelligence*, pages 31–40, 2006.

[7] G. Governatori, A. Rotolo, and G. Sartor. Temporalised normative positions in defeasible logic. In *Proc. ICAIL '05*, pages 25–34, 2005.

[8] A. J. I. Jones and M. Sergot. A formal characterisation of institutionalised power. *Journal of the IGPL*, 3:427–443, 1996.

[9] D. Makinson and L. van der Torre. What is input/output logic? input/output logic, constraints, permissions. In *Dagstuhl Seminar Proc. 07122*, 2007.

[10] N. Oren, S. Panagiotidi, J. Vazquez-Salceda, S. Modgil, M. Luck, and S. Miles. Towards a formalisation of electronic contracting environments. In *Proc. of Coordination, Organization, Institutions and Norms in Agent Systems, the Int. Workshop at AAAI 2008*, pages 61–68, 2008.