

The AISB 2010 Symposium on Social Networks and Multiagent Systems

One of the most interesting research topics in the field of multiagent systems is the definition of models with the aim of representing social structures such as organizations and coalitions, to control the emergent behavior of open systems. Organizations and coalitions are composed by individuals, related to each other by different possible kinds of relations such as dependencies on goals, conflicts on resources, similar beliefs and so on. One important issue is how to represent these relations. Moreover, like human organizations, these social structures are characterized also by an high degree of dynamism.

In dealing with societal issues, the multiagent systems field took inspiration mostly from organizational theory in economics and legal theory, while less attention is devoted to the research area describing the relations among the individuals inside human organizations and their dynamics: social network analysis. Social network analysis has emerged as a key technique in modern sociology, anthropology, social psychology, communication studies, information science, organizational studies, economics as well as a popular topic of study. Social network analysis views social relationships in terms of nodes and ties. Nodes are the individual actors within the networks, and ties are the relationships between the actors. Research in a number of academic fields has shown that social networks play a critical role in determining the way problems are solved, organizations are run, and the degree to which individuals succeed in achieving their goals.

Despite the common object of study, multiagent systems and social network analysis use concepts like agents, relationships, dependencies, and so on which often have only superficial similarities. The aim of this symposium is to underline the differences and the similarity points between social network analysis and multiagent systems in the representation of the social structures and their dynamics, and to promote the interchange of knowledge and methodologies among the two research fields.

The first edition of the Symposium was in 2009 at the AISB Convention with the theme “Adaptive and Emergent Behaviour and Complex Systems”. Both social networks analysis and multiagent systems deal with the behavior of complex systems composed by many agents and have among their research topics adaptivity and emergence, since it is not realistic to consider only top-down rigidly designed social systems. This first edition highlighted that the two disciplines of multiagent systems and social network analysis need each other in order to define models with the aim of representing social structures such as organizations and control the emergent behavior of open systems, as underlined also by the Agentlink RoadMap (www.agentlink.org).

The interest in the papers of this second edition of the Symposium goes from social network analysis concerning the social network Facebook showing how social interactions among users come out from the grouping of common interests to the application of finite state machines and fuzzy set methods to the simulation of the behaviour of multiagent systems modeled on the notion of power. A relevant issue in this edition consists in the presence of papers about norms in multiagent systems and social network analysis. This could be an almost unexplored and fascinating proposal of bridging the two areas.

Giulia Andrighetto, Guido Boella, Ugo Pagallo and Serena Villata

Programme Chairs:

Giulia Andrighetto, ISTC-CNR Rome
Guido Boella, University of Turin
Ugo Pagallo, University of Turin
Serena Villata, University of Turin

Programme Committee:

Patrice CAIRE, University of Luxembourg, LUXEMBOURG
Cristiano CASTELFRANCHI, ISTC-CNR Rome, ITALY
Rosaria CONTE, ISTC-CNR Rome, ITALY
Terrill FRANTZ, Carnegie Mellon University, USA
Francesca GIARDINI, ISTC-CNR Rome, ITALY
Gustavo A. Gimenez LUGO, Federal Technological University of Parana, BRASIL
Brian HIRSHMAN, ISR Computation, Organizations and Society, USA
Daniel T. MAXWELL, Innovative Decisions Inc., USA
Jeremy PITT, Imperial College London, UK
Luigi SAURO, University of Naples, ITALY
Jaime SICHMAN, University of Sao Paulo, BRASIL
Carles SIERRA, IIIA-CSIC, SPAIN
Pietro TERNA, University of Turin, ITALY
Leendert VAN DER TORRE, University of Luxembourg, LUXEMBOURG
Harko VERHAGEN, Stockholm University, SWEDEN

What is Power ?

Perspectives from sociology, multi-agents systems and network analysis.

Matthias Mailliard¹ and Christophe Sibertin-Blanc²

Abstract. Although power is a core and well debated concept in sociology and political sciences, its definitions remain fuzzy and partial in both the fields of multi agents systems (MAS) and social network analysis (SNA). *Power on, power of, influence, coercion, centrality* : What is *power* ? We propose a complex perspective on *power* principally based on the French sociology: *the strategic analysis*. We argue that *power* has a transactional nature based on the action triptych: *to give, to receive, to concede*, and expose some limits and extents about the use of *power* in sociological, MAS, and SNA organizational analysis.

1 INTRODUCTION

Multi-agents systems (MAS) and social network analysis (SNA) are cross-disciplinary fields, kinds of pivots exchanging concepts between a large range of scientific areas. This two fields also produce tools to formalize, to measure, and to understand various objects. The problem, that can be an advantage, is that that cross-disciplinary concepts are often suffering from mutations due to their reification. It is especially the case with social sciences. Sometimes it is a positive evolution of a term, like the minimalist definition of agent proposed by Ferber [1] that meets an agreement in many disciplines. Sometimes it is damageable because the formalization reduction only focuses on parts of the addressed concept or problem. This is what we want to develop here concerning the sociological notion of power.

On this topic, the long time relationship between social sciences and SNA, has certainly advantaged this later. Indeed, from Moreno [2], to Molm [3], Freeman[4] or Grosseti [5], passing by Granovetter [6], Blau [7] or Burt [8], there is, since a few decades, a very strong bond between these disciplines. This dynamic has begun to tie up in the USA, and it is grounded in the sociology of organizations and social exchange theories.

We continue this section by exposing the, maybe, common root of all these works (in sociology of organization, MAS, and SNA) on power: Max Weber [9], [10]. Then we develop power perspective in social exchange theories and, after that, how social exchange theories has been developed in MAS.

We are not sure that lineage: exchange theory (Homans [11], Blau [7]) → power-dependence theory (Emerson [12]) → network dependence theory in MAS ([13], [14], [15], [16]), is the only and complete way to catch the concept of power, even if it is very fruitful.

So, we propose to climb down the tree until Blau and Crozier meeting and to explore another branch of the sociology of organizations: the French *sociology of organized action* also known as *strategic analysis* ([17], [18], [19]). We introduce and formalize it in section 2.

Section 3 introduces several indices aiming at seizing power in a complex perspective and providing many points of view on power: dyadic, systemic, relative, strategic - that can be interpreted in many theories as strategic analysis or gift-giving paradigm [20].

We conclude by showing the limits of our work, and propose a way to combine it with other measures of power from SNA.

1.1 The sociological roots

So, what is the starting point? As a part of each political ideology, power, its forms, its expression, and its management, has been thought as far as human beings had to think the organization of the city, the society. We thus don't pretend here, and are not legitimate, to cover authors as Platon, Machiavelli, Hobbes, Marx or Rousseau, who have deeply spoken on power. We modestly want to start our argumentation with modern sociology, beginning with the political sociology of Weber, because : (1) "every organizational approaches found their roots in the work of Weber" [21], (2) as we will show, MAS and SNA are (partially) founded on sociology of organizations and actively take part to its construction, and (3) these three research fields address more or less directly the notion of power.

Lawyer, historian, economist and sociologist Max Weber (1864-1920) is considered to be one of the founders of modern sociology. As such, its work has been well studied, and one can refer for example to [9] or [10].

In its studies of social phenomena, Weber gives the primacy of explanation of the actors' actions, in a posture known as *methodological individualism*. That is, he tries to explain the structure and the evolution of some social facts by questioning the actors' motivations, and, in a somewhat empathic way, by understanding what are the causalities of their actions.

In the field of political sociology, he establishes a *theory of legitimate domination*, where the *domination* is a special social relation between actors, where one is *ordering* another and the other one orients his behavior toward this order according to its *validity*. The validity is assessed depending on the order *legitimacy*, and it will likely determine a conforming behavior, an avoidance behavior or a disobedience one. Following the methodological individualism posture, Weber propose four typical reasons for a subject to legitimate an order: the tradition, the affective or emotional belief, the value-rational belief (*value-rationality*) and the legality ; one can also follow an order only on the base of a pure rational computation (*instrumental rationality*). Reasoning on salient features of phenomena, *ideal types*, Weber thus propose three types of domination : the *charismatic domination*, the *traditional domination*, and the *rational-legal domination* which characterize the *bureaucratic organization* of the modern state. He then exposes the routinization of organizations based on charismatic domination into a traditional or a rational-legal one. The modern times

¹IRIT – University of Toulouse, France, matthias.mailliard@gmail.com

²IRIT – University of Toulouse, France, sibertin@univ-tlse1.fr

We are thankful to our "reviewers": Paul Chapron and Samuel Thiriot.

routinization is thus based on sciences and legal rationality at the detriment of the magical, the mystical or the religious, what Weber calls “disenchantment of the world”.

To resume for our purpose, Weber gave us a notion of domination as an imbalanced relation; structuring organizations; whose maintain and evolution are ensured by rational motivated behaviors of actors. He provides a definition of domination as “*Macht bedeutet jede Chance, innerhalb einer sozialen Beziehung den eigenen Willen auch gegen Widerstreben durchzusetzen, gleichviel worauf diese Chance beruht.*”³ [22].

Following Weber, many works, in many disciplines, have been done on the organizational phenomenon, and at the very beginning they were rather focused on the mechanist theories of the contemporaries, Taylor [23] and Fayol [24], than on the sociology of Weber. As a simplification, we can say that these sociological approaches about organizations not only focus on firms; they adopt the frame of a theory of action (for example as T. Parson or R.K. Merton) which considers the organization not as an entity but as the product of the activity of social actors provided with cognitive abilities, goals that are their own, and are constantly making choices (instead of being essentially reactive, according to the school of human relations). They always integrate a view of the organization and adopt the principles of bounded rationality of Herbert Simon [25][26] (individuals are not omniscient and adopt the first solution found that lead them to a better situation than the current one) and procedural rationality (i.e. a decision is taken not because it is the best in terms of outcomes, but because the decision procedure is considered as the best available).

1.2 Power and social exchange

The issue of social exchange, it will be addressed by exchange theories or in the paradigm of gift-giving [19] (or *potlatch* [27]), has led to develop interesting concepts of power. Moreover, we will show later that authors about social exchange have also inspired many works in social network analysis and in agent-based social simulation.

1.2.1 Exchange theories

Exchange theories are mainly founded on the works of George Homans [11] and Peter Blau [7], the former more interested in the structuring of interpersonal relationships at a micro-level (groups), and the later rather in the complex structuring of macro-sociological phenomena. The theory then develops with Emerson [12], which proposes a reticular and well formalized conception of exchange. The network theory of exchange assumes that power is based on resources, access to alternative sources, frequency of interactions and connections between relations. Linda Molm summarizes and develops the theory [28] by including the dimension of coercive power [3] with an interesting reasoning on its exercise and the risk of breaking the relationship an abuse could entail⁴.

1.2.2 Gift and potlatch

³It seems there is a debate on the translation in English. Here is the translation proposed by Talcott Parsons: “*The probability that one actor within a social relationship will be in position to carry out his own will despite resistance, regardless of the basis on which probability rest.*”

⁴About this kind of process we can also refer to Hirschman [29] who expose a twofold ways to resist when loyal: voice (i.e. explicitly entering in confrontation) and exit (i.e. breaking the relation).

Potlatch is basically a ceremony in the course of which goods are distributed among rival tribes in North America during the 19th century. The more the gift offered to another was important, the more this one was obliged to reciprocity or overbidding, and had to get rid of this debt under the menace to enter in conflict, or, even, in war. This particular form of exchange was erected by Marcel Mauss as a model of gift-exchange, based on a threefold obligation *to give, to receive and to reciprocate*. It is also known as the gift-giving paradigm.

Marcel Mauss [19] has raised the potlatch as a major concept of anthropology : a *total social fact*, that is to say with: religious, morphological, legal, shamanic, aesthetic and economic features.

Due to its nature of being a total social fact, the potlatch seems furthermore interesting to understanding of power: the threefold obligation to give, to receive and to reciprocate; the waste of goods as a way to express a superior position of domination; symbolic exchanges; creation and maintenance of social links; acquisition of prestige, status and social roles; etc.

1.3 Social exchange in MAS

The availability of already formalized theories lets an easier introduction of *the exchange theories* in the field of MAS and social simulation. These works are mainly founded, among others, on the work of Castelfranchi ([13], [14]), Sichman [15], Conte [16]. These authors summed up the results of the network exchange theory to produce frameworks for the representation of dynamic and open social organizations. By this way, the agents of these organizations are able to represent themselves the organizational system, and to mobilize this knowledge into decision making processes. This is for example suggested in [15], [30], [31], [32]. Social mechanisms of coordination are then proposed on the basis of these representations, as such, the engagement of agents in coalitions [15], depending on their goals and some measurement of their dependence that they compute from their knowledges or believes.

In these works, power is the counterpart of the *dependence*. It subsumes different concepts: *power to* (autonomy of action and access to resources), *power on* (an actor that is dependent to complete a goal), or the *institutional power* [33].

The MAS approach of exchange theories can in fine be viewed as a descriptive theory, more useful for (agent) decision making, than for understanding social phenomena. It is thus not a coincidence that these works are mainly used in the context of software applications such as peer-to-peer systems or auctions. If they have led to a comprehensive approach of mechanisms involving trust and reputation, we regret that the same effort was not oriented for mechanisms involving/of power, especially as the sociological exchange theories are closely interested in, especially, the work of Linda Molm about coercive power [28].

It is somewhat surprising that potlatch, which, according to [34], has greatly inspired the founders of exchange theories, has not successfully spread in MAS. However, two works in agent-based simulation deal with the gift-giving paradigm:

- The work of J. Rouchier and F. Bousquet [35] studies the construction of the reputation of individuals within a society based on trade in terms of *prestige gifts* or *share gifts*, following their desire of prestige or social integration.
- The work of [36], about an artificial market of tasks, introduces the possibility of gift exchanges to build long term dyadic relationships. They are supposed to enable alliances, to reinforce confidence, and to organize market.

2 FORMALIZING STRATEGIC ANALYSIS

2.1 Sociology of organized action

Crozier and Friedberg have developed a sociological theory of organizations, indifferently called *sociology of organized action* or *strategic analysis*, whose main objective is to unveil actors' context of action to understand why they behave as they do. They are interested in organizations more or less structured whose members are engaged in middle or long term relationships for the achievement of the organization's purpose, including organizations such as administrations, firms, institutions, etc, whose boundaries and action context are more or less well defined.

Crozier and Friedberg, define organizations as *concrete systems of action*, structured as and by interleaved games of power. Power is an imbalanced relation where actors exchange behaviors. The exercise of power is exerted by means of whatever resources, called *zones of uncertainties*. They are the ingredients of action, they mediate the interactions, and thus are the support of power relations.

Relations of power are multidimensional, n-ary, interleaved. We thus propose to decompose them into chunks, called *relations*. Each one include a (*relational*) *state* attribute whose value is *controlled* by only one actor and corresponds to the specific behavior he adopts. This behavior is constrained by his *room for manœuvre* and will determine the access by others to the resource underlying the relation. This access is more or less valuable for the actors following their resulting *action capability*, and the amount of interest, or *stake*, they accord in accessing this resource. Finally, the state of a relation can affect the *room for manœuvre* of another relation.

The actors are considered as bounded rational and their actions are goal oriented: in a methodological perspective (*methodological utilitarianism*) they are considered as being strategic; or in a Weber or Boudon style, they always have good reasons to do what they do. The actors' strategic rationality make them to increase or maintain their control on the resources of power (i.e. *zones of uncertainties*) in order to get a high action capability from others. Their actions are constrained by the power relation structure which in turn is affected by their actions.

2.2 Formalization of strategic analysis

On this base, we have developed a formal meta-model of organizations *à la* Crozier and Friedberg [50]. We expose here the main part of this formalization:

Relations

- R is the set of *Relations*.
- $control \subset R \times A$ denotes that an actor controls a relation. The $a \rightarrow r$ is equivalent to $a \text{ control } r$.
- R_a is the set of relations *controlled* by an actor a .
- $S_r = [-1; 1]$ denotes the space of states for a *relation* r .
- $M_r = [b_{min_r}; b_{max_r}]$ represents the space of states of a *relation* r that can be accessed by its controller, that is its *room for manœuvre*. It is characterized by its boundaries : $-1 \leq b_{min_r} \leq b_{max_r} \leq 1$.
- $S_R = \prod_{r \in R} S_r$ is the set of state vectors of the relational structure. $s_R \in S_R$ is called a *situation*.

- $S_O = \prod_{r \in R} S_r^3$ is the set of the states of the organization.
- $constrain \subset R \times R$ denotes that the *room for manœuvre* of a *relation* is determined by another one. We assume that the notation $r_1 \rightarrow r_2$ is equivalent to $r_1 \text{ constrain } r_2$. We treat cases where if $r_1 \rightarrow r_2$ then $\exists r_3, (r_3 \rightarrow r_1) \vee (r_3 \rightarrow r_2)$.
- If $r_1 \rightarrow r_2$, $M_{(r_1, r_2)}: E_{r_1} \rightarrow E_{r_2} \times E_{r_2}$ is the function which determines the *room for manœuvre* of the relation r_2 .

Actors

- A is the set of actors.
- $(E_{action_cap}, +, \cdot) = (\mathbb{R}, +, \cdot)$, is the ordered vectorial space of action capabilities. We can thus combined elements together or with a real (the *stake*, the *solidarity*).
- $E_{payoffs} = [-10; 10] \subset E_{action_cap}$ is the payoffs space.
- $E_{stake} = [0; capital_{stake}] \subset \mathbb{R}$ is the set of stakes such that $capital_{stake} = 10$.
- $E_{solidarity} = [-capital_{solidarity}; capital_{solidarity}]$ is the domain of solidarities such that $capital_{solidarity} = 1$.
- $Actions_r = [-2; 2]$ is the action set for a relation r . An action a is valid for a relation state e_r if it is in the scope of the *room for manœuvre*: $e_r + a \in M_r$.
- $Action_a = \prod_{r \in R_a} Actions_r$ is the action set for an actor a . It is composed by the action sets of each relation r such as $a \text{ control } r$.
- $Effect_r: E_r \times A \rightarrow E_{payoffs}$ is the function that fixes the action capability that an actor can mobilized using the resource underlying the relation r . It depends on the state of r , i.e. the behavior the controller of r adopts. The result of the effect function of r is called the *payoffs* given by r .
- $stake: A \times R \rightarrow [0; 10]$, denotes the relative importance an actor gives to a relation to fulfill its goal. The amount of stake. We assume that $a \text{ depends on } r$, $r \rightarrow a$ or $stake(a, r) > 0$ are equivalent.
- $solidarity: A \times A \rightarrow [-1; 1]$ sets the positive or negative solidarity of one actor toward another. Negative solidarity can express conflict, animosity, or concurrent goals, while positive solidarity expresses the usual definition of the terms, friendship, or a necessary structure for a successful cooperation. We assume that the amount of absolute solidarity is the same for each actor $\forall a \in A, \sum_{b \in A} |solidarity(a, b)| = 1$. Solidarity is aimed at orienting action of actor by helping or confronting others.

2.3 Organization as a social game

As demonstrated in [37], and following Crozier and Friedberg, a concrete system of action can be seen as a social game. This game is running, embedding as many models of rationality as we have hypotheses about how power transactions occur, evolve, and regulate the organization into a stabilized structure of exchanged behaviors.

3 COMPLEX MEASUREMENT OF POWER

Thanks to the mathematical structure of organization, inspired from the Crozier and Friedberg theory, we are able to define and

compute quantities that are relevant from many paradigms, such as gift-giving, cost/advantage and strategic analysis. This gives us a complex perspective on the concept of power.

3.1 Impact of the controller behavior on the action capability of the dependent actors

Impact is an index that measures the action capability a relation provides to an actor depending on its state. The *impact* of a relation represents its contribution to what is globally received by an actor, and that we called his *satisfaction*.

Impact of a relation on an actor is a combination of:

- the *effect* of the controller behavior, that is the relation's *state*, on the actor's action capability; it is the *payoffs* given by a relation;
- the *stake* of the actor for this relation.

Moreover, an actor can take into account the impact of the relation on others, following the nature and the intensity of its links of solidarity.

We thus have two ways to compute *impact* of a relation r on an actor a for a given relation state s_r :

- If we only focus on actors payoff and stake, we will speak of *ego-centered impact*, noted: $impact_{ego}$:

$$impact_{ego}(r, a, s_r) = effect(r, a, s_r).stake(r, a) \quad (1)$$

- If we take its solidarities into account we will speak of *socio-centered impact*, noted $impact$ as its the canonical form of impact we use:

$$impact(r, a, s_r) = \sum_{b \in A} sol(a, b).impact_{ego}(r, b, s_r) \quad (2)$$

As impact is the base of other indices, there is always an *ego-centered* and a *socio-centered* version for each index.

3.1 Satisfaction: to receive

By estimating and combining the impacts of each relation he depends on, an actor has the ability to assess the situation, i.e. what he receives from others. We call this combination his *satisfaction*. For an observer, *satisfaction* is the whole measure of the action capability an actor can mobilize to fulfill its goals, and eventually to counterattack or to cooperate in the fulfillment of others' goals. For an actor, *satisfaction* is a way to estimate his current situation, to relate it to previous decisions and to make the next one. It is thus of first importance to assess the quality of transactions in the course of power relationships. A basic way to compute satisfaction in a situation s_R , is to add-up impacts:

$$satisfaction_{ego}(a, s_R) = \sum_{r \in R} impact_{ego}(r, a, s_r) \quad (3)$$

$$satisfaction(a, s_R) = \sum_{r \in R} impact(r, a, s_r) \quad (4)$$

It is also of interest to know the amount of satisfaction an actor provides to himself by the way of the relations he controls:

$$controlled\ satisfaction(a, s_R) = \sum_{r \in R_a} impact(r, a, s_r) \quad (5)$$

and to know the overall satisfaction of the organization. This index is a tool to guide the analysis of simulations or to detect big incoherences in organization design, but it is clearly not a measure of the vitality, or the well being inside an organization:

$$cumulated\ satisfaction(s_R) = \sum_{a \in A} satisfaction(a, s_R) \quad (6)$$

3.2 Power: to give

The *power*, as defined here, is an index of the exercised power. Following the thesis of Crozier and Friedberg, this index reveals a part of the “imbalanced exchange of behaviors”. It expresses what is given to others in terms of action capabilities during, or at the end, of the exchange process, and thus it depends on the state of the involved relations. This measure takes into account the subjectivity of actors (*stake*) and their solidarities, it is thus based on *impact*. *Power* can be seen as the amount of positive or negative satisfactions an actor provides to one or many others.

If we adopt the definition of power given by R. Dahl (quoted in [19]): “*A has power over B to the extent that he can get B to do something that B would not otherwise do*”, then the index we propose returns “*the means implemented by A to get B to do Y*” (without guarantee of success). In this perspective, a positive contribution of A to the satisfaction of B can be interpreted as:

- a gift forcing B to reciprocity (to do Y)
- a counter-gift to honor a debt contracted by A to B
- the fact of yielding to a pressure exerted by B
- an opportunity, due to transactions of A with other players, that benefits B.

Conversely, a negative contribution of A to the satisfaction of B can be interpreted as:

- a reaction to the non-reciprocity of B in its trade with A,
- a coercive means to constrain B to do Y
- a response to a pressure of B
- a penalty, due to transactions of A with other players, that penalizes B.

Accordingly, we define the *power* of a over b , *relative power* of a over b , *effective power*, or *cumulated power* for a given actor:

$$power(a, b, s_R) = \sum_{r \in R_a} |impact(r, b, s_r)| \quad (7)$$

$$relative\ power(a, b, s_R) = power(a, b, s_R) - power(b, a, s_R) \quad (8)$$

$$cumulated\ power(a, s_R) = \sum_{b \in A, b \neq a} power(a, b, s_R) \quad (9)$$

These indices are underlying postulates showing that, for Crozier and Friedberg, power is expressed as the way an actor knows how to be unpredictable. By using absolute *impacts* in equation (7), we thus evaluate the intensity of power for one actor, that is how things can change, for one (equation (7)) or all other actors (equation (9)), depending on his behavior. Actor's power of high intensity can be the result of high positive and negative impacts on distinct relations, and so could be considered as inconsequential in other(s)' goal fulfillment. Nevertheless, he is thus armed with high room for manœuvre, making him unpredictable^{5,6}.

A more effective version of power should also be considered as it expresses the amount of satisfaction given by an actor to another one (X) or everyone else (X):

$$eff\ power(a, b, s_R) = \sum_{r \in R_a} impact(r, b, s_r) \quad (10)$$

⁵The interpretation can differ to what we just say, and have to be validated by correlation with more observations (the number of involved relations, the previously played actions, the fact that some relations could be blocked ...).

$$\text{cumulated eff power}(a, s_R) = \sum_{b \in A, b \neq a} \text{eff power}(a, b, s_R) \quad (11)$$

The SocLab⁷ environment [40] shows these indices (and many others) in the form of matrices or as labels tied to edges or vertices of the organizations' network, so the cited indices of *power* can be rapidly compared to each other⁸.

3.3 Concession: to concede

The notion of concession appeared as soon as we have addressed the construction of a measure of the strategic use of power, following the *methodological utilitarianism* and the strategic nature of actors [19]. We remind that we are not in a one move game: in social game, the end is, for each actor, to find a behavior that is satisfactory for himself and for others, so that the organization can go on in this way. This sustainable behavior is what we call his strategy.

Whether the equation is based on: "I'm as strategic as I give little and receive plenty", so the equation becomes easy: *strategic* = *receive* – *give* = *satisfaction* – *power*. But what is given is not what is lost⁹.

The giving of action capabilities through *relations* also has an effect on the actor's own capabilities. He is ready to concede (that is, literally, to cede in a relationship) something because he expects a (positive) social reaction (reciprocity, obedience, ...). We thus propose to measure what an actor concedes, by comparing how he could have satisfied himself by its own, let's call this *autonomy*, with how he effectively satisfies himself, let's call this *controlled satisfaction*.

3.3.1 Autonomy

Let's first reason on only one relation r . Taking into account potential constraints on r , we can compute the best impact the controller of r can obtain only by himself, in order to maximise his satisfaction (this is his Nash-equilibrium behavior). Being careful about the cases where he is also the controller of r' such as $r \rightarrow r'$ or $r' \rightarrow r$, we can compute *autonomy* as:

$$\text{autonomy}(a) = \sum_{r \in R_a} \max \text{certain impact}(r) \quad (12)^{10}$$

3.3.2 Concession

We are now able to propose two concession indices: a dyadic one, and a systemic one. Concession of on actor a to an actor b is the difference, regarding the relations he controls and b depends on, between what he could have and what he actually has:

$$\begin{aligned} \text{concession_to}(a, b, s_R) \\ = \sum_{r \in R_{a \rightarrow b}} \max \text{certain impact}(r) - \sum_{r \in R_{b \rightarrow a}} \text{impact}(r, a, s_R) \end{aligned} \quad (13)$$

We can assess, in a dyad, the difference between the concession of each actor. This *relative concession* is, in some way, the dual of the *relative power*. *Relative power* is founded

on a difference of gifts, while *relative concession* is founded on a difference of losses. The first index enlightens gift-giving logic of power, the one who has power is the one who gives the most and proceeds to the creation of a debt that will take its full potential over time, while the second enlightens a cost / advantages logic:

$$\begin{aligned} \text{relative concession}(a, b, s_R) \\ = \text{concession}(a, b, s_R) - \text{concession}(b, a, s_R) \end{aligned} \quad (14)$$

In the systemic version, concession is not computed by adding the dyadic version for each actors:

$$\begin{aligned} \text{concession}(a, s_R) \\ = \text{autonomie}(a) - \text{controlled satisfaction}(a, s_R) \end{aligned} \quad (15)$$

It is because *relations* can impact many actors, so, if for example $a \rightarrow r$, $r \rightarrow b$, and $r \rightarrow c$, then what a concedes to b is the same value as when a concedes to c . So we cannot distinguish to which actor, b or c , a is conceding.

This reveals a problem in many sociological attempts to define power, as the ones of Crozier or Dahl¹¹. These definitions consider power as having a dyadic nature, although it has a system wide extent.

The counterpart of this not dyadic nature is that actors can use opportunistically *concession_to* as an argument in negotiation with several actors at once.

3.4 Transactional advantage

Going one step further, it is possible to propose a measure of the negotiation quality of an actor a , involved in transactions with b , by counting the difference between what b gives to a and what a concedes to b , and by removing the minimum that b must give to a . This gives a measure of the strategic exercise of power of an actor, we call *transactional advantage*. The dyadic version is given as follows:

$$\begin{aligned} \text{transactional_advantage}_{(a,b,s_R)} \\ = \text{power}(b, a, s_R) \\ - \min_{s_R}(\text{power}(b, a, s_R)) \\ - \text{concession_to}(a, b, s_R) \end{aligned} \quad (16)$$

This definition suffers from the same shortcomings as indicators of dyadic concession so we will prefer the systemic version, *global transactional advantage*, based on the systemic version of the *concession*:

$$\begin{aligned} \text{transactional_avantage}_{\text{global}}(a, s_R) \\ = (\sum_{b \in A, b \neq a} \text{power}(b, a, s_R)) - \text{concession}(a, s_R) \end{aligned} \quad (17)$$

4 CONCLUSIONS

To support our considerations about power, we have propose an index toolkit that permits to construct one's own definition of power, following the considered organization and topic. We have exposed major issues in sociological and network exchange theory definitions of power:

- considering power in a dyadic way lead to approximations: power is systemic,

¹¹I set apart their definitions from their overall works that covers far more than we discuss here.

⁶A good way to measure the potential unpredictability of an actor a is to measure its *structural power*:

$$\text{structural_power}(a) = \max(\text{power}(a)) - \min(\text{power}(a))$$

⁷<http://sourceforge.net/projects/soclab/>

⁸We thus don't need to use a *mean power* $((\text{power}(a, b) + \text{power}(b, a)) / 2)$ to seize the whole intensity of power between two actors, as in [38].

⁹Indeed, social games are not zero sum games (while some economic games are).

¹⁰*max certain impact* is a measure of the maximum impact an actor can obtain by his own. It is too long to be exposed in the present paper.

• in social exchange, what is given must be think in conjunction with what is conceded.
Considering these two points, the strategic exercise of power can be evaluated accurately.

Agent-based social simulation could benefit of the proposed indices:

- analysis of organization: they enable a precise and multi-level analysis of power games and could be complete with social network analysis (see next subsection),
- analysis of organization design: they enable to validate the global coherence of organizational model. Usage in mechanism design must be considered (see last section).
- agent rationality: indices can be used in the decision making process of strategic agent ([39], [49]), or as arguments in negotiation process. Imprecision of dyadic indices make them good candidates for bound rational agent ([39], [49]).

4.1 Articulating MAS and SNA perspectives on power in sociological analysis

The proposed multi-agents formalization of *concrete systems of action* and our indices are expressive enough to handle the main types of resource of power. Due to the amount of informations needed to feed such models, they are dedicated to organizations with a small number of (individual or collective) actors ([40], [41]). We thus introduce the possibility of a smart joint use with SNA. First, SNA have some indices good at handling communicative type resources of power, such as *betweenness centrality* [4] or *closeness centrality* [42]. Second, these indices don't need large amount of informations, so that they can be used on large scale representations of social organizations. For these reasons, we think that there are two possible ways to articulate our proposal with SNA:

- *Strategic zoom*: a SNA where we zoom in a subset of the network to make a strategic analysis.
- *Reticular zoom*: a strategic analysis where we zoom in a collective actor to make a SNA analysis [43].

4.2 Mechanism design

Other works could be of interest to our ones, especially *mechanism design*. *Mechanism design* is often used in economics, political science or theory of votes to define rules ensuring a fair operating, no abuse from ones at the expense of others. Mechanism design is at the crossroads of social choice and equity theories ([44],[45],[46]), game theory, and has reached computer sciences and MAS ([47], [48]). These works are particularly interested in coalitions, their formations, and power.

5 REFERENCES

[1] Ferber, J., 1999. *Multi-Agent Systems*. Addison-Wesley.
[2] Moreno, J. L., 1986. *Théâtre de la spontanéité*.
[3] Molm L. 1997. *Coercive power in social exchange*. CU Press
[4] Freeman, L. C., 1977. A set of measures of centrality based on betweenness. *Sociometry*, 40:35{41.
[5] Grossetti, M. 2004. *Sociologie de l'imprévisible*, PUF.
[6] Granovetter, M., 1978. *Threshold Models of Collective Behavior*. The American Journal of Sociology, Vol. 83, No. 6. (May, 1978).
[7] Blau P. 1964. *Exchange and Power in Social Life*, J. Wiley & Sons.

[8] Burt R. 1992. *Structural holes. The Social Structure of Competition*, Cambridge, Harvard University Press, 310 p.
[9] Aron, R., 1976. *Les étapes de la pensée sociologique*, Gallimard.
[10] Boudon R., 2005. *Pouvoir*. In *Dictionnaire de la pensée sociologique*. PUF, Paris.
[11] Homans G. 1958. *Social Behavior as Exchange*. Am. Jo. Sociology.
[12] Emerson R. 1976. *Social Exchange Theory*. An. Rev. Sociology.
[13] Castelfranchi C. 1990. *Social Power: a missed point*, in DAI, MA and HCI. In *Decentralized AI*. Y. Demazeau & J.P.Mueller.
[14] Castelfranchi C., Cesta A., Miceli M., *Dependence relations among autonomous agents*. Decentralized A.I.3: Elsevier., 1992.
[15] Sichman, J., 1994. *Du Raisonnement Social Chez les Agents: Une Approche Fondée sur la Théorie de la Dependence*. PhD Thesis.
[16] Conte, R., Sichman, J., 1995. *DEPNET: How to benefit from social dependence*. Journal of Mathematical Sociology. Vol.20. 2--3.
[17] Crozier, M. *Le phénomène bureaucratique*, Paris, 1963.
[18] Crozier, M., Friedberg, E., 1977. *L'acteur et le système*, Seuil, Paris
[19] Friebberg, E., 1993. *Le Pouvoir et la Règle*, Seuil, Paris.
[20] Mauss, M., 1924. *Essai sur le don. Forme et raison de l'échange dans les sociétés archaïques*. Année Sociologique, 1923-1924
[21] Crozier, M., 2005. *Organisation (Théorie des)*, Dictionnaire de la pensée sociologique, PUF, Paris.
[22] Weber, M. (1971 [1921]) *Économie et société*, Paris, Plon.
[23] Taylor, F.W., 1902. *Shop Management*, A.S.M.E., vol. XXIV, 1902.
[24] Fayol, H., 1916. *Administration industrielle et générale*, Dunod.
[25] Simon H., Newell A., 1972. *Human problem solving*, Prentice-Hall.
[26] Simon, H., A., 1969. *The Sciences of The Artificial*.
[27] Malinowski B.. 1922. *Argonauts of the Western Pacific*. Dutton.
[28] Molm, L., 1997. *Coercive Power in Social Exchange*, CU Press.
[29] Hirschman, A., 1970. *Exit, Voice and Loyalty*, Harvard Univ. Press.
[30] Lopez y Lopez F, Luck, M., *Empowered Situations of Autonomous Agents*, in *Advances in Artificial Intelligence IBERAMIA* 2002.
[31] Hübner J., Sichman J. S., Boissier O., 2004. *Using the MOISE+ for a Cooperative Framework of MAS Reorganization*.
[32] Herzig & al., *Prolegomena for a logic of trust and reputation*. NorMAS'08, Luxembourg, 2008.
[33] Jones A, Sergot M, 1996. *A formal characterisation of institutionalised power*. Journal of the IGPL.
[34] Cook K., Mauzet M. , 2005. *Echange social*. In *Dictionnaire de la pensée sociologique*, PUF, Paris.
[35] Rouchier J., Bousquet F., O'Connor M., 2001, *Emergence of reputation in an artificial society*, JASSS, vol 4 issue 2.
[36] Alam S. J., Hillebrandt F. and Schillo M. (2005) *Sociological Implications of Gift Exchange in Multiagent Systems*, JASSS, v.8.3.
[37] Mailliard M., & al., *Cooperation is not always so simple to learn....* AESCS IV. Springer-Verlag, p. 147-154, september 2007.
[38] Molm L., 1990. *Structure, Action, and outcomes: the dynamics of power in social exchange*. American Sociological Review.
[39] Mailliard, M., Roggero, P., Sibertin-Blanc, C., 2006. *Un modèle de la rationalité limitée des acteurs sociaux*, JFSMA'06, Hermès.
[40] Mailliard M. 2008. *Formalisation Multi-Agents de la Sociologie de l'Action Organisée*. PhD Thesis, Toulouse, december 2008.
[41] Vautier C., Roggero P., Adreit F., Sibertin-Blanc C., 2009. *Evaluation by simulation of the social acceptability of agricultural policies for water quality*, 4th Int. ICSC Symposium on Information Tech. for Environment Engineering, ITEE'09, Springer-Verlag.
[42] Sabidussi, G., 1966. The centrality index of a graph. *Psychometrika*.
[43] Sturma, A., 2007. *Expertise et interdisciplinarité*. Master Thesis.
[44] Arrow K., 1963. *Social Choice and Individual Values*. J. Wiley & S.
[45] Coleman J., 1990. *Foundations of Social Theory*, Harvard U. Press.
[46] Sen A., 1979. *Equality of What?* Stanford University.
[47] Rosenschein & Zlotkin, 1994. *Rules of Encounter*, MIT Press.
[48] Dash R.K., Parkes D., Jennings, N.R., 2003. *Computational Mechanism Design : A Call to Arms*. IEEE Intelligent Syst. 18(6)
[49] Chapron P., El Gemayel J., Sibertin-Blanc C., 2009. *Impact of tenacity upon the Behaviors of Social Actors*. Applied Agent based simulator Engineering for Complex System study, Nagoya, Japan.
[50] Sibertin-Blanc C., Amblard F., Mailliard M., 2006. *A Coordination Framework based on the Sociology of Organized Action*. COIN'06.

Optimizing the Core Computation with Social Networks

Luigi Sauro¹ and Serena Villata²

Abstract. Cooperative boolean games are a family of coalitional games where agents may depend on each other for the satisfaction of their personal goals. In this work we investigate two different types of social networks, *abstract dependence networks* and *refined dependence networks*, that are used to define the notion of *stable coalitions* and Δ -reduction, respectively. Stable coalitions enable to focus on a subset of the agents and use results to determinate the *core* of the whole game. Δ -reduction prunes the search space by returning a set of actions that are not admissible to be executed. We present an algorithm based on stable coalitions and a Δ -reduction implemented in Prolog and experimental results that show how they effectively improve the computation of the core.

1 Introduction

Cooperative Boolean Games [5] are a new particular family of boolean games [7, 6]. In such type of games, agent's primary aim is to achieve its individual goal, which is represented as a propositional logic formula over some set of boolean variables. Each agent is assumed to exercise a unique control over some subset of the overall set of boolean variables, and the set of truth assignments for these variables corresponds to the set of actions the agent can take. As secondary aim, each agent wants to minimize the cost associated to the execution of its actions. As in typical coalitional games, an agent can have the necessity to cooperate with the others because it does not have sufficient control to ensure its goals are satisfied. However, the desire to minimize costs leads to preferences over possible coalitions, and hence to strategic behavior. One of the solution concepts proposed for such games is the notion of core that strengthens the Nash equilibrium: a truth assignment of the boolean variables (strategy profile) is in the core in case no subset of agents can unilaterally deviate from it and improves the rewards of each of its members.

In this paper, we propose a new step to make the computation of the core easier by means of the social dependence networks associated to the cooperative boolean game. The reasons of our choice are twofold: first, we present a number of abstractions that allow to reduce the search space by means of a set of criteria principally based on graphs visit algorithms which are computationally tractable; second, we underline a number of hidden properties in the notion of core showing how, in certain cases, this notion is too much strict and, thus, it can lead to counterintuitive results.

We define two kinds of social dependence networks, representing two different levels of abstraction of a cooperative boolean game. Abstract dependence networks have already been used by Bonzon et al. [2] to define stable coalitions which enable to divide the entire problem in subproblems and then combine their solutions. While

Bonzon et al. [2] show that the notion of stability is complete with respect to the pure Nash equilibrium with non costly actions, we show in this paper that the notion of stability is complete also with respect to the solution concept of the core in case of cooperative boolean games with costly actions. In Bonzon et al. [3], the authors extend their previous results with a generalization to non-dichotomous preferences, where agents can not only express plain satisfaction or plain dissatisfaction, but also intermediate levels. This generalization suffices to replace the preference component of a boolean game by an input expressed in a propositional language for compact preference representation.

After Castelfranchi [4], Boella et al. [1] and Sauro [9] show how to use social dependence networks to discriminate among different potential coalitions during the coalition formation process. They develop a criterion of admissibility called *do-ut-des* property describing a condition of reciprocity: an agent gives a goal only if this fact enables it to obtain, directly or indirectly, the satisfaction of one of its own goals. This criteria has only a qualitative connotation, it cannot be directly applied to the solutions developed in game theory. In this approach goals are not structured and they do not represent explicitly the costs of the actions.

Refined dependence networks essentially provide a graph representation of a cooperative boolean game where the numerical information about costs is abstracted and actions are simply partitioned in free and costly actions. We present a reduction, called Δ -reduction, to pass from a cooperative boolean game G to a cooperative boolean game (CBG) G' , simpler to be solved because less actions can be executed. In Sauro et al. [10], the introduction of abstract and refined dependence networks is provided. We extend the results presented in [10] with the definition of the algorithm allowing the application of the Δ -reduction and the experimental results it allows to achieve.

The reminder of the paper is as follows. In section 2 we formally define cooperative boolean games. Section 3 shows the optimization techniques relative to the core-membership problem. Sections 4 and 5 present the algorithms and the experimental results. Conclusions end the paper.

2 Cooperative Boolean games

A Cooperative Boolean Game (CBG) [5] consists of a set of agents $1, \dots, n$ which desire to accomplish personal goals. Goals are represented by propositional formulas γ_i over some set of boolean variables Φ . Each agent controls a (possibly empty) subset $\Phi_i \subseteq \Phi$ of the overall set of boolean variables. The notion of control is used to mean that agent i has the unique ability within the game to set the value (either $\top$ or $\perp$) of each variable $p \in \Phi_i$. Variables are assumed to be initially false, this way of setting a variable $p \in \Phi$ to be $\top$ has the meaning of *performing the action* p , whereas setting $p \in \Phi$ to $\perp$ means *doing nothing*. Since action (as opposed to inaction) typi-

¹ Dipartimento di Scienze Fisiche, University of Naples

² Dipartimento di Informatica, University of Turin

cally incurs some cost, Dunne et al. [5] define the cost function $cost$ in such a way that $cost(p)$ denotes the cost of performing action p (i.e., setting p to $\top$). Agents' secondary goal is to minimize its costs. Summing up briefly, if the only way an agent can achieve its goal is by making all its variables true, then the agent prefers to to this rather than not achieve its goal but if there are different ways to achieve it, then the agent prefers always those that minimize costs.

Definition 1 A cooperative boolean game G is a $(2n + 3)$ -tuple

$$G = \langle A, \Phi, cost, \gamma_1, \dots, \gamma_n, \Phi_1, \dots, \Phi_n \rangle$$

where $A = \{1, \dots, n\}$ is a set of agents, $\Phi = \{p, q, \dots\}$ is a finite set of boolean variables, $cost$ is a cost function defined in $\Phi \rightarrow \mathbb{R}_+$, $\gamma_1, \dots, \gamma_n$ are the boolean formulas over Φ which represent the goals of the agents and $\Phi_1, \dots, \Phi_n$ is a partition of Φ over n , with the intended interpretation that Φ_i is the set of Boolean variables under the control of agent i .

A subset ξ of Φ is a valuation, where the usual meaning is that the value of the variables belonging to ξ is true and the value of the other ones is false. $\xi \models \phi$ means that ϕ is true under the valuation ξ under the standard propositional semantics.

Valuations intuitively correspond to possible strategies of the agents, in Dunne et al. [5], $cost_i(\xi)$ denotes the cost to agent i of valuation $\xi \subseteq \Phi$, that is,

$$cost_i(\xi) = \sum_{v \in (\xi \cap \Phi_i)} cost(v)$$

As in Dunne et al. [5], we define an utility function that is always positive if the valuation ξ satisfies the goal of the agent i and otherwise it is always negative. However, this is not an utility function in the classical way as discussed by von Neumann and Morgenstern [8] such as, for example, a function $u : X \rightarrow \mathbb{R}$ from choices to the real numbers assigning a real number to every outcome in a way that captures the agent's preferences over both simple and compound lotteries. In [5], if μ represents the total cost of all variables, the utility for agent i of a valuation ξ , $u_i(\xi)$ is defined as:

$$u_i(\xi) = \begin{cases} 1 + \mu - cost_i(\xi) & \text{if } \xi \models \gamma_i \\ -cost_i(\xi) & \text{otherwise} \end{cases}$$

This utility function leads to a preference order $\succeq_i$ over valuations: $\xi_1 \succeq_i \xi_2$ iff $u_i(\xi_1) \geq u_i(\xi_2)$. The best utility for an agent is reached if it has its goal satisfied without performing any action (utility of $\mu + 1$) while the worst utility for an agent is reached in the case it does not get its goal satisfied but it performs actions such as to set all its variables true (utility $-cost_i(\Phi_i)$).

Typically, an agent has to cooperate with the other agents because it does not have sufficient control to ensure its goal is satisfied. However, aiming at minimizing its costs, an agent is preferentially unwilling to execute one of the actions under its control. Therefore, agents cooperate only in the case in which a cooperative solution is preferable to the alternatives, either because the agents cannot achieve their goals independently or their can reduce their respective cost. The utility achieved by agents within a coalition depends not just on their actions but on the actions of the whole set of agents involved in the game.

A coalition D is represented as a (sub)set of agents, $D \subseteq A$ and thus it does not represent any kind of particular relationship among the members composing it. Let $D \subseteq A$ be a coalition, then Φ_D denotes the set of variables under the control of some member of D ,

$\Phi_D = \bigcup_{i \in D} \Phi_i$. If a valuation ξ_2 is the same as a valuation ξ_1 except at most in the value of the variables controlled by D then $\xi_1 = \xi_2 \text{ mod } D$.

D blocks ξ_1 through ξ_2 since ξ_2 could do better than ξ_1 only by flipping the value of some of the variables under the control of D . From Dunne et al. [5], the definition of blocked valuation is as follows:

Definition 2 (Blocked Valuation)

A valuation ξ_1 is blocked by a coalition $D \subseteq A$ through a valuation ξ_2 if and only if:

1. ξ_2 is a feasible objection by coalition D which means that $\xi_2 = \xi_1 \text{ mod } D$.
2. D strictly prefers ξ_2 over ξ_1 : $\forall i \in D : \xi_2 \succ_i \xi_1$.

The blocked valuation allows us to define the core of a cooperative boolean game. The core is a fundamental concept behind coalitional game theory. A valuation is in the core if and only if no coalition has an incentive to defect.

Definition 3 (Core)

Given a cooperative boolean game G , $\xi \in \text{core}(G)$ if and only if it is not blocked by any coalition.

We focus now on the problem of computing the core of a CBG. In general, this problem can be seen as a generation and test problem: generate a strategy, then check whether it belongs to the core according to Definition 3. Thus, it is useful in the following to consider a *generation space*, that is the set of strategies ξ we need to check core-membership, and a *test space*, the set of strategies ξ' we search to establish whether ξ belongs to the core. Clearly, if no optimization is implemented both the generation and the test space consist in the set of all strategies and unfavorable have a cardinality equal to $2^{|\Phi|}$.

3 Reducing the generation space

In Sauro et al. [10], we introduce two optimization techniques that enable to reduce the generation space. Both of them are based on the notion of social dependence network that can be defined starting from a CBG. These social networks, on the one hand, explicitly represent the inter-dependencies among agents according to their goals, on the other hand they abstract from the quantitative aspects of a game associated to the cost function. In particular, the first type of social dependence network, we named Abstract Dependence Networks, is the same already used in Bonzon et al. [2] as decomposition method for the Nash Equilibrium in Boolean Games without costly actions. In Sauro et al. [10] we essentially extend the results in Bonzon et al. [2] to the framework defined in Dunne et al. [5].

As in Bonzon et al. [2], in order to correctly establish the dependencies among agents we need to define which variables are relevant for the satisfaction of the agents' goal. A variable p is said irrelevant for a formula ϕ in case there exists an equivalent formula ϕ' where p does not occur. With $RV_G(i)$ we represent the set of all variables $p \in \Phi$ that are relevant for γ_i , whereas $RA_G(i)$, is the set of agents $j \in A$ such that j controls at least one relevant variable of i . The notions of relevant agent and relevant variable are crucial in what follows since they allow to define dependence networks and the way to prune them. Using the notion of relevant agents, we define a dependence network where nodes represent agents and an edge from i to j represents the dependence of i on j ($j \in RA_G(i)$).

Definition 4 (Abstract Dependence Network)

Given the CBG $G = \langle A, \Phi, cost, \gamma_1, \dots, \gamma_n, \Phi_1, \dots, \Phi_n \rangle$, the abstract dependence network of G is the directed graph $ADN(G) = \langle N, R \rangle$ such that the set of nodes N correspond to the agents in G , $N = A$, and $(i, j) \in R$ iff $j \in RA_G(i)$.

As in Bonzon et al. [2] we say that a set of agents in $ADN(G)$ is stable in case it is closed under the relation R . $R(C)$ is the set of players from which C may need some action in order to be satisfied.

Definition 5 (Stable set) Given a directed graph $\langle N, R \rangle$, $C \subseteq N$ is stable iff $R(C) \subseteq C$, i.e. for all $i \in C$, for all j such that $(i, j) \in R$ $j \in C$.

Definition 6 (ADN Projection)

Let $G = \langle A, \Phi, cost, \gamma_1, \dots, \gamma_n, \Phi_1, \dots, \Phi_n \rangle$ be a CBG, $ADN(G) = \langle N, R \rangle$ the corresponding abstract dependence graph and $C = \{i_1, \dots, i_m\} \subseteq N$ a stable set, the projection of G on C is defined by $G_C = \langle C, \Phi_C, cost_C, \gamma_{i_1}, \dots, \gamma_{i_m}, \Phi_{i_1}, \dots, \Phi_{i_m} \rangle$, where $cost_C : \Phi_C \rightarrow \mathbb{R}^+$ is the restriction of cost on the set of boolean variable of C , Φ_C .

By definition the goals occurring in the projection of a CBG on a stable set C does not contain variables controlled by agents outside C , therefore the projection is itself a CBG.

By using stable sets, Abstract Dependence Networks can be safely used to split the original problem in subproblems without loosing solutions. However, Abstract Dependence Networks may hide some useful information that can also be used to prune some strategies that cannot belong to the core - and hence to reduce the search space. For this reason we define another notion of dependence network at a lower level of abstraction called Refined Dependence Networks (RDNs). These social networks may seem actually equivalent to the boolean game itself, except for the cost of the variables. These costs, however, are not a minor point since two boolean games that result in the same RDN can have different solutions.

A Refined Dependence Network represents how the goals can be satisfied by means of AND-arcs among the agents whose single edges are labeled with literals. Furthermore, costly actions are marked in a set Δ .

Definition 7 (Refined Dependence Network)

A Refined Dependence Network is a AND-graph $\langle N, \Phi, \Delta, E, \Phi_1, \dots, \Phi_n \rangle$ where N is the set of nodes, Φ is the set of boolean variables, $\Delta \subseteq \Phi$ is the subset of costly variables³, $E \subseteq N \times 2^{(N \times \text{Litt}(\Phi))}$ where $\text{Litt}(\Phi) = \Phi \cup \{\neg p \mid p \in \Phi\}$ and $\Phi_i \subseteq \Phi$ where n is the cardinality of N .

Given a literal l , we denote by $|l|$ the corresponding boolean variable. If $l \in \Phi$, then $|l| = l$ whereas if $l = \neg p$ then $|l| = p$. Furthermore, to simplify the formalism, we represent an AND-arc $(i, \{(j_1, l_1), \dots, (j_m, l_m)\})$ as the set of triples $\{(i, j_1, l_1), \dots, (i, j_m, l_m)\}$. Of course, a set as $\{(1, 3, p), (3, 4, q)\}$ has no meaning in our context.

We use Refined Dependence Networks to reveal the structure of interdependencies among agents. First, we assume that the goals of the agents do not contain irrelevant variables and are given in disjunctive normal form, i.e. $\gamma_i = \gamma_{i_1} \vee \dots \vee \gamma_{i_m}$ where each γ_{i_j} is a conjunction of literals [11]. To simplify again the formalism we describe respectively γ_i as a set of γ_{i_j} and each γ_{i_j} as a set of literals - the empty set has the usual meaning respectively of $\perp$ referred

³ The set of variables with an associated cost.

to γ_i and $\top$ referred to the γ_{i_j} . Roughly, each AND-arc e outgoing the agent i corresponds to a $\gamma_{i_j} \in \gamma_i$, where each single edge that composes e is labeled with a literal occurring in γ_{i_j} and reaches the agent that controls the corresponding variable. The set Δ consists of the actions that have a strictly positive cost.

Definition 8 (From CBGs to RDNs)

Given the CBG $G = \langle A, \Phi, c, \gamma_1, \dots, \gamma_n, \Phi_1, \dots, \Phi_n \rangle$, $RDN(G) = \langle N, \Phi, \Delta, E, \Phi_1, \dots, \Phi_n \rangle$ is such that $N = A$, $\Delta = \{p \in \Phi \mid c(p) > 0\}$ and $\{(i, j_1, l_1), \dots, (i, j_m, l_m)\} \in E$ iff $\{l_1, \dots, l_m\} \in \gamma_i$ and for all $1 \leq h \leq m$, $|l_h| \in \Phi_{j_h}$.

Example 1 Let G be a cooperative boolean game defined by $A = \{1, 2, 3, 4\}$, $\Phi = \{a, b, c, d, e\}$, $cost(a) = cost(b) = cost(c) = cost(d) = cost(e) = 1$, $\gamma_1 = a$, $\gamma_2 = c \wedge e$, $\gamma_3 = b \wedge c$, $\gamma_4 = d$, $\Phi_1 = \{b, e\}$, $\Phi_2 = \{d\}$, $\Phi_3 = \{a\}$, $\Phi_4 = \{c\}$. The associated refined dependence network $RDN_G = \langle A, \Phi, E, \Phi_1, \dots, \Phi_n \rangle$, where E is composed by the following dependencies: $\{(1, 3, a)\}$, $\{(3, 1, b), (3, 4, c)\}$, $\{(2, 1, e), (2, 4, c)\}$, $\{(4, 2, d)\}$.

Given a Refined Dependence Network $RDN(G) = \langle N, \Phi, \Delta, E, \Phi_1, \dots, \Phi_n \rangle$, we mean with $R_E \subseteq N \times N$ the binary relation such that $(i, j) \in R_E$ just in the case there exists an AND-arc $e \in E$ that starts from i and reaches j , i.e. for some literal l , $(i, j, l) \in e$. It is easy to see that $ADN(G) = \langle N, R_E \rangle$ and hence $RDN(G)$ describes G at a lower level of abstraction with respect to $ADN(G)$.

We want to use Refined Dependence Networks to impose some constraints to the set $core(G)$. To this scope some preliminary results are needed. A boolean variable $a \in \Phi_i$ is said to be *unfavorable* if and only if $a \in \Delta$, i.e. $cost(a) > 0$, and for each $\{l_1, \dots, l_m\} \in \gamma_i$, $a \notin \{l_1, \dots, l_m\}$. In the following we denote by $[i]^-$ the set of unfavorable variables of the agent i .

Proposition 1 Given a cooperative boolean game G and an agent $i \in A$, for each $a \in [i]^-$, $\xi \in core(G)$ implies $a \notin \xi$.

Proof 1 $a \in [i]^-$ means that $cost(a) > 0$ and a does not occur (positive) in γ_i . Assume that $\xi \models \gamma_i$, then, for some $\{l_1, \dots, l_m\} \in \gamma_i$, $\xi \models \{l_1, \dots, l_m\}$. Since $a \notin \{l_1, \dots, l_m\}$, this means that also $\xi \setminus \{a\} \models \{l_1, \dots, l_m\}$.

Now it remains to show that if for each ξ , $\xi \models \gamma_i$ implies $\xi \setminus \{a\} \models \gamma_i$, then $\xi \in core(G)$ implies $a \notin \xi$. Assume that $a \in \xi$, clearly $\xi \setminus \{a\} = \xi \text{ mod } \{i\}$ as $a \in \Phi_i$. Furthermore, as $cost_i(\xi \setminus \{a\}) < cost_i(\xi)$ and by hypothesis $\xi \models \gamma_i$ implies $\xi \setminus \{a\} \models \gamma_i$, then $\xi \prec_i \xi \setminus \{a\}$. But this means that ξ is blocked by i through $\xi \setminus \{a\}$.

Note that, according to Proposition 1, if all the variables are unfavorable, the core can contain only the empty strategy. We prove that a goal depending on an unfavorable variable a can be reduced into one that does not depend on a without affecting the possible solutions. More precisely, we define the notion of reduction as follows.

Definition 9 Given a cooperative boolean game G and an unfavorable variable $a \in [i]^-$, we say that the cooperative boolean game G' is a Δ -reduction of G just in the case it is obtained from G applying the following steps:

1. remove a from Φ_i ;
2. remove from each γ_i any conjunction of type $\{l_1, \dots, a, \dots, l_n\}$;
3. replace in each γ_i any conjunction of type $\{l_1, \dots, \neg a, \dots, l_n\}$ with $\{l_1, \dots, l_n\}$.

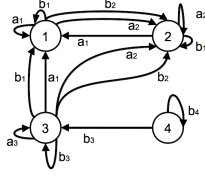


Figure 1. Refined Dependence Network of Example 1.

Proposition 2 Let G' be a Δ -reduction of a CBG G , $core(G) \subseteq core(G')$.

Proof 2 For each valuation ξ that does not contain the unfavorable variable a , it clearly holds for each agent i that $u_i(\xi)$ in G is equal to $u_i(\xi)$ in G' . Therefore, if in G' , ξ'_1 is blocked by C through ξ'_2 , then the same holds in G and hence $core(G) \subseteq core(G')$.

The converse does not hold. Note that the previous results do not use *quantitative* values of the cost function but only the fact that an action has a strictly positive value, therefore they reside in the level of abstraction of RDNs. We can now define a procedure on $RDN(G)$ which uses unfavorable variables and Δ -reductions to reduce the search space in finding the core.

Definition 10 (Reduction rule) Let $RDN(G)$ be a Refined Dependence Network and let ω denoting a valuation initially set to $\emptyset$, the reduction rule $RDN(G)$ is given by applying exhaustively the following rule:

Condition: for some $a \in \Phi_i \cap \Delta$, there does not exist an AND-arc e outgoing from i such that $(i, j, a) \in e$ (i.e. a is unfavorable).

Action: remove any AND-arc e' such that $(j, i, a) \in e'$, a from Φ_i and add a to ω .

Due to propositions 1 and 2, the valuation ω we obtain from definition 10 constraints the strategies in $core(G)$ to be also in $\Phi \setminus \omega$. For more details and graphical examples about ADNs and RDNs, see [10].

4 Reducing the test space

In the previous section, we have seen how to determinate the set of unfeasible actions ω such that the generation space becomes the powerset of $\Phi \setminus \omega$. However, even if core-membership has to be checked just for one strategy (for example the *empty* strategy $\emptyset \subseteq \Phi \setminus \omega$), the test space still remains the set of all strategies $\xi' \subseteq \Phi$, which is exponentially large in the size of a CBG. For this reason, in this section we study optimization techniques that reduce the test space.

Before studying new optimization techniques, it is reasonable to look for previous results that can be reused for our end. In particular, two properties shown in Dunne et al. [5] can be taken into account. Let $contrib(\xi)$ be the set of agents that incur some positive cost in ξ ($contrib(\xi) = \{i \mid \exists a \in \xi \text{ s.t. } a \in \Phi_i \text{ and } cost(a) > 0\}$) and $ben(\xi)$ the beneficiaries of ξ ($ben(\xi) = \{i \mid \xi \models \gamma_i\}$). Furthermore, we write $\xi \subset_C \xi'$ if $\xi \cap \Phi_C \subset \xi' \cap \Phi_C$ and say that ξ is C minimal for ϕ in case $\xi \models \phi$ and no $\xi' \subset_C \xi$, $\xi' \models \phi$. Then, the following properties hold:

1. $\xi \in core(G) \implies contrib(\xi) \subseteq ben(\xi)$

2. $\xi \in core(G) \implies \xi$ is $contrib(\xi)$ minimal for $\gamma_{contrib(\xi)}$

Each ξ in the generation space such that either $contrib(\xi) \not\subseteq ben(\xi)$ or it is not $contrib(\xi)$ minimal for $\gamma_{contrib(\xi)}$ can be discarded without checking whether it is blocked, thus, both the properties attempt to reduce the test space to the empty set. Whether they can be profitably used as optimization technique clearly depends on the computational complexity.

Since the condition $contrib(\xi) \not\subseteq ben(\xi)$ can be easily computed, the first property is a possible optimization technique (actually it has been implemented in our framework). On the contrary, with a proof that is essentially the same as Theorem 1 in Dunne et al. [5], checking that ξ is not $contrib(\xi)$ minimal for $\gamma_{contrib(\xi)}$ has the same computational complexity as checking that ξ belongs to the core (coNp-complete). Therefore, it has not been considered as a possible candidate.

Note that, since $contrib(\emptyset) = \emptyset$, the empty strategy cannot be discarded according to the first property for at least one strategy we still have a test space of exponential size. For this reason, we consider some other criterion to reduce the test space. Clearly, as in the case of unfeasible actions, we need a computationally easy way to select a subset of 2^Φ that is complete with respect to the core-membership problem.

On the one hand, when a strategy ξ is optimal for an agent i , i.e. $\xi = \argmax_{\bar{\xi}} u_i(\bar{\xi})$, no other strategy can be strictly more profitable for i , and hence i does not have any incentive to participate to a coalition D in order to block ξ . On the other hand, assume that ξ satisfies the goal γ_i , and for a given $a \in \Phi_i$, $cost(a) > cost_i(\xi)$, then i has no incentive to perform a . The following theorem expresses these two considerations, the proof is straightforward and it is left to the reader.

Proposition 3 Given a CBG G , a strategy ξ and an agent i . Assume that D blocks ξ through ξ' . We have that:

1. $\xi = \argmax_{\bar{\xi}} u_i(\bar{\xi}) \implies i \notin D$;
2. $\xi \models \gamma_i$, $a \in \Phi_i$ and $cost(a) > cost_i(\xi) \implies a \notin \xi'$.

Proposition 3 can be used to reduce the test space as follows. If ξ is optimal for the agents $O = \{i_1, \dots, i_m\}$, then a strategy ξ' that blocks ξ is such that $\xi \cap \Phi_j = \xi' \cap \Phi_j$, for all $j \in O$. Thus, the test space is reduced to $\bigcup_{h \notin O} \Phi_h$. Furthermore, if $cost(a) > cost_i(\xi)$, the test space can be further reduced to those strategies ξ' such that $a \notin \xi'$.

Finally, note that to decide whether a strategy ξ is optimal for an agent i it is not required to check for all the other strategies $\bar{\xi}$ with $u_i(\xi) \geq u_i(\bar{\xi})$. As we are considering goals in disjunctive normal form, i.e. $\gamma_i = \{\gamma_{i_1}, \dots, \gamma_{i_m}\}$, we just have to consider the strategies $\xi_{i_j} = \{a \in \Phi \mid a \in \gamma_{i_j}\}$. Then, the maximal value u_i^m of the utility $u_i(\xi_{i_j})$, with $1 \leq j \leq m$, corresponds to the maximal utility agent i can obtain. Therefore, we simply have to check that $u_i(\xi) = u_i^m$.

5 Algorithms

We use the results in Sections 3 and 4 to implement in Prolog a procedure `find_core` that, given as input a CBG G , returns $core(G)$. A procedural description of `find_core` is given in Algorithm 1. Essentially, `find_core` consists of two procedures, `FIND_CORE` and `CORE_MEM`.

In `FIND_CORE`, the ADN of G is instantiated and it is partitioned in its smallest subgraphs $A = \{A_1, \dots, A_n\}$ that are pairwise disconnected. Clearly, each A_i represents a stable coalition and no agent

occurs in two distinct A_i and A_j (lines 2-3). For each $A \in \mathbf{A}$, the projection G' of A is computed (line 6) and the core of G' is determined as follows. First, the RDN of G' is instantiated and, according to Definition 10, DELTA_RED calculates the set ω of unfeasible actions. Then, for each ξ not containing actions in ω , the core-membership of ξ is decided by CORE_MEM. If $\xi \in \text{core}(G')$, then it is added to CORE' .

The core of G is computed by gathering any union of strategies in each $\text{core}(G')$. This is done incrementally in line 14. In CORE_MEM, first it is checked whether $\text{contrib}(\xi) \subseteq \text{ben}(\xi)$, in affirmative case ξ cannot belong to the core and hence false is returned. CORE_MEM computes in line 4 the set O of agents such that ξ is optimal. According to Proposition 3, such agents do not have any incentive in modifying their strategies. Thus, their strategies are *frozen* in ξ_O . Then, given the actions TA of the remaining agents, actions a of the beneficiary agents with $\text{cost}(a) > \text{cost}(\xi)$ are discarded – again according to Proposition 3.

```

Input: A cooperative boolean game
       $G = \langle A, \Phi, \text{cost}, \gamma_1, \dots, \gamma_n, \Phi_1, \dots, \Phi_n \rangle$ 
Output:  $\text{core}(G)$ 
1  $\text{CORE} \leftarrow \{\emptyset\};$ 
2  $\text{ADN} \leftarrow \text{CONVERT\_ADN}(G);$ 
3  $\mathbf{A} \leftarrow \text{PAIRWISE\_DIS}(\text{ADN});$ 
4 forall  $A \in \mathbf{A}$  do
5    $\text{CORE}' \leftarrow \emptyset;$ 
6    $G' \leftarrow \text{PROJECT\_CBG}(G, A);$ 
7    $\text{RDN} \leftarrow \text{CONVERT\_RDN}(G');$ 
8    $\omega \leftarrow \text{DELTA\_RED}(\text{RDN});$ 
9   forall  $\xi \subseteq \Phi[G'] \setminus \omega$  do
10    if  $\text{CORE\_MEM}(\xi, G')$  then
11       $\text{CORE}' \leftarrow \text{CORE}' \cup \{\xi\}$ 
12    end
13  end
14  $\text{CORE} \leftarrow \{\xi \mid \xi = \xi_1 \cup \xi_2 \text{ where } \xi_1 \in \text{CORE} \text{ and } \xi_2 \in \text{CORE}'\}$ 
15 end
16 return  $\text{CORE};$ 

```

Algorithm 1: FIND_CORE

```

Input: A cooperative boolean game  $G$  and a strategy  $\xi$ 
Output: True if  $\xi \in \text{core}(G)$ , false otherwise.
1 if  $\text{contrib}(\xi) \not\subseteq \text{ben}(\xi)$  then
2   return false;
3 end
4  $O \leftarrow \{i \mid \xi = \arg\max_{\xi} u_i(\xi)\};$ 
5  $\xi_O \leftarrow \xi \cap \bigcup_{i \in O} \Phi_i;$ 
6  $TA \leftarrow \bigcup_{j \notin O} \Phi_j;$ 
7 forall  $j \in \text{ben}(\xi) \setminus O$  do
8    $TA \leftarrow TA \setminus \{a \mid a \in \Phi_j \text{ and } \text{cost}(a) > \text{cost}_j(\xi)\};$ 
9 end
10 forall  $\xi' \subseteq TA$  do
11   if  $\text{BLOCKED}(\xi, \xi_O \cup \xi')$  then
12     return false;
13   end
14 end
15 return true;

```

Algorithm 2: CORE_MEM

After this last optimization, the core membership is *brutally* computed and a strategy blocking ξ is searched in the set of strategies $\xi_O \cup \xi'$, where ξ' is a subset of the resulting TA .

As noted in Bonzon et al. [2], a further optimization could be in principle possible by removing some arc (i, j) in $\text{ADN}(G)$ in case all the actions of j that occur in γ_i are irrelevant. However, selecting irrelevant variables a in a formula ϕ means to check the validity of

$\phi_{\top} \leftrightarrow \phi_{\perp}$, where $\phi_{\top}$ and $\phi_{\perp}$ are obtained by substituting each occurrence of a in ϕ with $\top$ and $\perp$ respectively. Thus, this would add a coNP-complete step just to deal with a few cases of *pathological* goals and without being sure that, by removing an edge, two sub-graphs result disconnected. For this reason $\text{ADN}(G)$ is instantiated from the initial goals, without determining irrelevant variables.

6 Experimental results

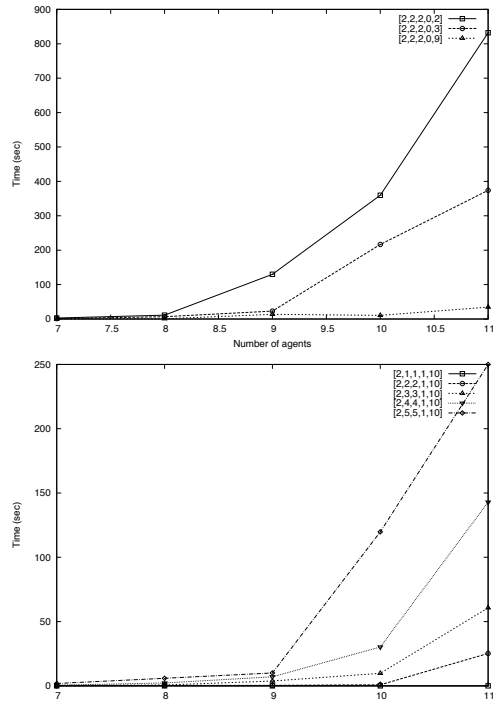


Figure 2. Experiments varying the costs of variables and the size of goals

We have implemented a generator that enables to instantiate CBGs of different *shapes*. The generator takes as input 6 parameters: the number of agents N_A , the number of actions per agent N_{Ac} , the number of disjuncts N_D , the number of conjuncts N_C , the minimal and maximal cost values C_{min} and C_{max} . As output we obtain a CBG with N_A agents that control N_{Ac} variables each. An agent's goal takes the form $\gamma_1 \vee \dots \vee \gamma_{N_D}$ where each γ_i consists in a conjunction $l_1 \wedge \dots \wedge l_{N_C}$ of randomly generated literals. Finally, an integer cost value from C_{min} to C_{max} is randomly assigned to each propositional variable. In the figures, the ordinate axis represents the run-time expressed in seconds where, for each input setting, the reported results correspond to the mean over 20 runs of `find_core`. In the figures the following tuple is used for the experiments $[N_{Ac}, N_D, N_C, C_{min}, C_{max}]$.

In Figure 2.a, the number of actions per agent as well as the structure of goals remain fixed ($N_{Ac} = N_D = N_C = 2$). Also, C_{min}

is fixed to 0 whereas the X axis corresponds to the number of agents N_A . Different lines correspond to different values of the maximal cost C_{max} . As can be seen, performances strongly depend on C_{max} and the framework behaves better by increasing it. This is only apparently surprising since costless actions cannot be discarded by applying Δ -reduction or in lines 7-9 of CORE-MEM. Now, as the cost of each action a is randomly chosen in the range $0, \dots, C_{max}$, the smaller C_{max} is the higher is the probability that $cost(a) = 0$. In particular, for $C_{max} = 2$ approximately 33% of the actions cannot be unfeasible, this rate decreases to 25% and 10% for $C_{max} = 3$ and $C_{max} = 9$, respectively. Furthermore, dealing with randomly generated CBGs, by setting $N_D = N_C = 2$ the resulting ADNs are very likely to be strongly connected and hence stable coalitions practically do not affect performances.

Figure 2.b shows how the framework behaves by increasing the size of goals. As before, $N_{Ac} = 2$ and the X axis corresponds the number of agents N_A , on the contrary the cost range is fixed to 1–10 – note that all variables are costly. Each line corresponds to different goal sizes, N_D is set equal to N_C and their value varies from 1 to 5. With goals composed by a single literal $N_D = N_C = 1$, even with 11 agents⁴, find_core returns in 0.2 seconds. By comparing with the previous figure, one of the factors that improves so much performances is that the resulting ADNs present quite often two or more disconnected components and hence stable coalitions actually decompose the original game.

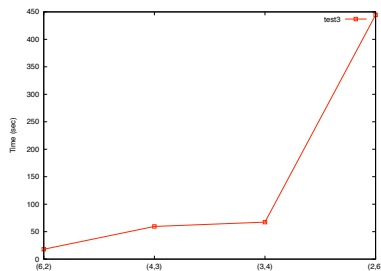


Figure 3. Varying the number of disjuncts and conjuncts in a goal

As said before, with $N_A = N_C = 2$ stable coalitions practically do not affect performances anymore. However, by generating random CBGs, goals are still enough *small* with respect to the number of possible literals that it is very likely that all – or all except – actions result to be unfeasible. So, in most of the cases, the find_core reduces the search space to the only empty strategy and hence basically this setting measures performances of CORE-MEM. For larger values of N_D and N_C , performances get worse. One reason is that the probability that a costly action is unfeasible decreases with the number of literals occurring in a goal. Thus, the Δ -reduction becomes less and less effective.

In Figure 2, bigger goals are obtained by increasing both N_D and N_C at the same time, but which of these two values has a higher impact on performances? Figure 3 shows a set of experiments where the number of agents, actions per agents and cost range is fixed (respectively $N_A = 7$, $N_{Ac} = 3$, $C_{min} = 1$ and $C_{max} = 10$). The

X axis represents different pairs of values for (N_D, N_C) such that the total number of literals composing a goal is constant, namely (6, 2), (4, 3), (3, 4), (2, 6). As before, we can see a valuable variability in performances and, to better understand why large N_C jeopardizes more than large N_D , we run a version of our framework with only Δ -reduction on values (6, 2) and (2, 6). The run-times obtained –respectively 457 and 449 seconds– are very close to the value obtained in the original framework for (2, 6). This means that the optimizations designed to decide the core membership perform better with goals composed by several disjuncts of small size than those with few disjuncts of big size.

7 Conclusion

In this paper we present an approach to optimize the computation of the core in cooperative boolean games [5] which is essentially based on dependence networks [9, 11]. The problem of computing the core of a cooperative boolean game G is a typical generation and test problem. We provide optimization techniques using two kinds of social dependence networks for the generation phase and extending some the results provided by Dunne et al. [5] for the test phase.

On the one hand, the advantage, shown by the results of Section 6, is that, also in the case of games with a number of strategies sized in an order of some millions, results are obtained in few minutes. On the other hand, the optimization techniques we propose in this paper are referred only to actions with a positive cost. Actions with no cost have not been analyzed in order to propose new optimization techniques and this is left for future research. Finally, the proposed optimization techniques, concerning the core-membership problem, return satisfactory results if they are applied to particular kinds of goals in which the disjuncts are small.

REFERENCES

- [1] G. Boella, L. Sauro, and L. van der Torre, ‘Strengthening admissible coalitions’, in *17th European Conference on Artificial Intelligence, ECAI*, pp. 195–199, (2006).
- [2] E. Bonzon, M. Lagasque-Schiex, and J. Lang, ‘Dependencies between players in boolean games’, in *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 9th European Conference, ECSQARU*, pp. 743–754, (2007).
- [3] E. Bonzon, M. Lagasque-Schiex, and J. Lang, ‘Dependencies between players in boolean games’, *Int. J. Approx. Reasoning*, **50**(6), 899–914, (2009).
- [4] C. Castelfranchi, ‘Social power. a point missed in multi-agent, dai and hci.’, in *Decentralized AI - Proceedings of MAAMAW’90*. Elsevier Science Publishers, (1990).
- [5] P. Dunne, W. van der Hoek, S. Kraus, and M. Wooldridge, ‘Cooperative boolean games’, in *7th International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, pp. 1015–1022, (2008).
- [6] P. Harrenstein, *Logic in Conflict*, Ph.D. dissertation, Utrecht University, 2004.
- [7] P. Harrenstein, W. van der Hoek, J. Meyer, and C. Witteveen, ‘Boolean games’, in *8th Conference on Theoretical Aspects of Rationality and Knowledge, TARK*, pp. 287–298, (2001).
- [8] J. Neumann and O. Morgenstern, *Theory of Games and Economic Behaviour*, Princeton University Press, 1944.
- [9] L. Sauro, *Formalizing admissibility criteria in coalition formation among goal directed agents*, Ph.D. dissertation, University of Turin, 2005.
- [10] L. Sauro, L. van der Torre, and S. Villata, ‘Dependency in cooperative boolean games’, in *Agent and Multi-Agent Systems: Technologies and Applications, 3rd KES International Symposium, KES-AMSTA*, volume 5559 of *Lecture Notes in Computer Science*, pp. 1–10, (2009).
- [11] J. Sichman and R. Conte, ‘Multi-agent dependence by dependence graphs’, in *1st International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS*, pp. 483–490, (2002).

⁴ Note that $N_A = 11$ and $N_{Ac} = 2$ mean $2^{22} \approx 4 \cdot 10^6$ possible strategies composing initially both the generation and the test spaces.

A Process Based on the Fuzzy Set Theory for Evaluation of Link Prediction Methods

Elisandra Aparecida Alves da Silva¹, Marco Túlio Carvalho de Andrade¹

Abstract. Link Prediction is an important research line in the Social Network Analysis context, as predicting the evolution of such nets is a useful mechanism to improve and encourage communication among users. This paper describes a process based on the Fuzzy Set Theory for evaluation of Link Prediction methods. In the data selection step, the use of fuzzy sensors is considered, the determination of new links is based on fuzzy compositions and the fuzzy ROC curve is used to analyse the results.

1 INTRODUCTION

Nowadays many databases are described as a linked collection of interrelated objects. The networks formed by such objects can be homogeneous, in which there is a single object type and link type or heterogeneous networks in which objects and links may be of multiple types. An example of homogeneous network is the WWW (World Wide Web), and examples of heterogeneous networks include co-authorship networks, which are used in this project.

The main goal of traditional data mining algorithms is to find patterns in a dataset characterized by a collection of independent instances of a single relation. However, the application of traditional statistical inference procedures which use independent instances can lead to inappropriate conclusions about data [11]. According to Liben-Nowell [14], a challenge in the Data Mining area is to deal with richly structured, heterogeneous data. This way, linking features among objects need to be used to improve the accuracy of predictive models [7]. Some of these features are mentioned by Davidsen [4]: the correlation between attributes of interconnected objects and the existence of links between objects which have similarity.

Link Mining refers to Data Mining techniques that explicitly consider the links in the development of predictive or descriptive models of interconnected data. The Link Mining tasks, according to the taxonomy presented by Getoor [7], are: object-related tasks (Ranking, Classification, Clustering and Object Identification), link-related tasks (Link Prediction) and graph-related tasks (Subgraph Detection, Graph Classification and Generative Models for Graphs). Link Mining is an emergent area that represents the intersection of different areas: Link Analysis, Web and Hypertext Mining, Relational Learning and Inductive Logic Programming, and Graph Mining.

The main goal of Link Prediction is to determine the existence of a link between two entities using object or link attributes. Link Prediction is useful in different application

fields, such as recommendation systems, detection of links not observed in terrorist networks, protein interaction networks, prediction of collaborations between scientists and Web hyperlinks prediction [23].

In related works ad hoc strategies are adopted to evaluate the results of predictions. This paper presents a systematic process to evaluate Link Prediction methods that is based on non dichotomic metrics for data selection, determination of new links and evaluation of results. The use of non dichotomic metrics makes it possible to use the specialist's knowledge and to adopt a perspective more similar to the human perception of the problem. The ROC curve is used to evaluate the results, but this step is under construction.

The paper is organized as follows: Section 2 presents an overview of the main methods of Link Prediction and Section 3 presents the proposed process. Finally, the application of the process and the results are presented.

2 LINK PREDICTION

The main goal of Link Prediction (LP) is to predict the existence of a link between two entities using features of objects and other observed links.

The basic approach of Link Prediction methods is the classification of all node pairs based on graph proximity measure. The link weight called $score(x, y)$ is assigned to each node pair x and y , and then a list is generated in decreasing order of score. Considering node x , $\Gamma(x)$ denotes a neighbor set of x in G_{collab} . Neighbors of x are the nodes which are directly connected with x .

Thus, these methods can be seen as the computation of proximity measure or similarity between nodes x and y , related to the topology of the network. In general, the methods derive from Graph Theory and Social Network Analysis and are designed to measure similarity between nodes. According to Liben-Nowell [14], these methods need to be modified for applications in different contexts.

Many approaches are based on the idea that the greater the number of common neighbors between two objects, the greater the chance of a link between x and y . Davidsen [4] and Jin [12] have proposed abstract models for network growth using this idea. These authors present the most direct idea of the application of Common Neighbors to Link Prediction. Newman [17] has used this measure in the context of collaborations network.

$$score(x, y) = |\Gamma(x) \cap \Gamma(y)| \quad (1)$$

¹ Dept. of Computer Engineering and Digital Systems, Univ. of São Paulo, Av. Prof. Luciano Gualberto, travessa 3, 158, SP, Brazil. Email: {elisandra.silva, marco.andrade}@poli.usp.br.

Adamic [1] used the proximity idea to verify the similarity between personal web pages. They consider common neighbors with lower degrees as more relevant, as follows:

$$\text{score}(x,y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(|\Gamma(z)|)} \quad (2)$$

Another approach, called Preferential Attachment, assumes that the probability of a new link involving x and y is proportional to the number of their neighbor's links. This measure is given as follows [2]:

$$\text{score}(x,y) = |\Gamma(x)| \times |\Gamma(y)| \quad (3)$$

The Link Prediction methods presented above are based on structural properties of networks and do not consider the connection weights between users. Murata [16] proposed some adjustments based on proximity measures to be used in online social networks. As user's personal information is not generally available in these networks, only the structural properties have been used. Additionally, the connection weight between x and y , $w(x,y)$, was defined as the number of encounters between x and y . A simple adaptation of Common Neighbors including link weights is presented in Murata [16]:

$$\text{score}(x,y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x,z) + w(y,z)}{2} \quad (4)$$

In this context, there are also approaches based on path analysis [13], [8] and on graph structure of networks [10]. Different approaches consider the application of probabilistic models [23] and similarity measurements between two objects [20], [22]. The main problems related to these approaches are the high complexity of the probabilistic models and the utilization of node attributes in similarity measurements that sometimes are not available in networks. Additionally, the link's information is not considered in these approaches.

The next section deals with the fuzzy approaches. As Murata [16], we assume the existence of link weights between users in a social network, but considering attributes of co-authorship networks.

3 PROPOSED APPROACH

The determined process involves the steps presented in Figure 1. In the data selection step, the use of fuzzy sensors is considered, the determination of new links is based on fuzzy compositions and the fuzzy ROC curve is used to analyse the results. This way, we have a process based on non dichotomic metrics in order to evaluate the methods of Link Prediction, what makes it possible to use the specialist's knowledge and to adopt a perspective more similar to the human perception of the problem.

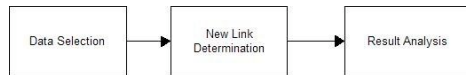


Figure 1. Evaluation Process

The following sections present the approaches used in the evaluation process.

3.1 Fuzzy Sensor

According to Mauris [15], the goal of a sensor is to generate a symbolic linguistic representation from numerical measurements, i.e., a numeric-linguistic conversion considering the subjectivity of the problem.

Thus, the sensor creates a symbolic qualitative description in two stages: (1) numeric measurement and (2) numeric-linguistic conversion. A numeric measure, generally obtained by an electronic processing, provides an objective quantitative description of objects. The measure language, usually obtained from the interrogation of users provides a qualitative description of subjective objects. The conversion should provide a very accurate description as the one performed directly by a human. Therefore, to implement a symbolic sensor, the symbolism of the adopted language should be considered in order to artificially reproduce the human perception of the measure.

The fuzzy sensors used for data selection includes two input variables: NumberOfPapers and NumberOfCoauthors and an output variable that determines the choice of the node. Input variables are named so because they represent authors of a co-authorship network. However, they can be obtained in different areas. The NumberOfPapers is the number of encounters with other users and the NumberOfCoauthors represents the number of neighbors.

3.2 New Link Determination

This work views the link weight between two users x and y as the "relation quality". This measure is obtained by the application of approaches that use features from co-authorship social networks, which can be directly used in other domains.

Two approaches based on fuzzy theory are presented here. Both consider the use of fuzzy compositions to determine new links between two authors and employ the relation quality to determine the link weight.

The approaches consider that the quality of the relation between two authors is higher in the following situations:

- When two authors have a large number of papers, mainly in recent years.
- When the average of co-authors of the authors in the relation is low, but the common co-authors are not considered as they influence the relation in a positive way.

The next sections present the technique used for new link prediction.

3.2.1 Fuzzy Compositions

In this paper, a new link prediction approach is proposed using fuzzy compositions. Supposing that $R(X,Y)$ and $R(Y,Z)$ are two fuzzy relations², the composition $Q(X,Z)$ between $R(X,Y)$ and

² A fuzzy relation establishes associations of different truth degrees between related elements, which are similar to the Fuzzy Set membership degrees [24]. A fuzzy relation example is given by "physical similarity between members from x and y ".

$S(Y, Z)$ is a fuzzy relation now between X and Z , using Y as a bridge (transitivity) [24]. It is given by:

$$C(X, Z) = R(X, Y) \circ S(Y, Z) \quad (5)$$

Figure 2 shows some user relations in a specific context as an example. It presents the link weights and their membership degrees.

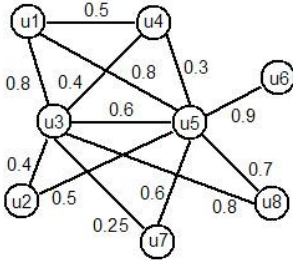


Figure 2. Users relations in a social network

Therefore, using relation compositions makes it possible to predict new link weights connecting users that have not been connected yet. The result of the composition to the previous network is presented in Table 1. The values in the example are obtained by using the Max-product. In the approaches, the Max-Product showed better results than the Max-Min.

Table 1. Relation Composition

	u3	u5		u7	u4	u1
u2	0.4	0.5	O	0.25	0.4	0.8
u8	0.8	0.7		0.6	0.3	0.8

	u7	u4	u1
u2	0.3	0.16	0.4
u8	0.42	0.32	0.64

3.2.2 Relation Quality

This measure represents the quality of the relation between two users. We propose two approaches to obtain this value. In the first approach this measure is the output variable of the fuzzy model and in the second, this is obtained by using a metric. The next subsections detail these approaches.

Using Fuzzy Model (Fuzzy1)

The input variables used in this approach are NumberOfPapers, CoauthorsAverage and RelationTime.

NumberOfPapers is the number of papers coauthored by A and B , considering different weights according to the paper's year. Recent papers are more important. The value a paper adds to the total number of papers is obtained as follows:

$$\text{PaperYear} - \text{InitialYearOfTraining} \quad (6)$$

Thus, the first year of the training period is not relevant to this measure and the total sum is the number of papers coauthored by A and B .

CoauthorsAverage is the average of coauthors of A and B , but the common coauthors are not considered. $\Gamma(A)$ is the number of coauthors of A and $\Gamma(B)$ is the number of coauthors of B . This value is obtained as follows:

$$C_o = \frac{\Gamma(A) + \Gamma(B)}{2} - (\Gamma(A) \cap \Gamma(B)) \quad (7)$$

RelationTime is the difference between the last year of training and the year the oldest paper was published.

The rules used are presented below:

if CoauthorsAverage is low AND NumberOfPapers is low AND RelationTime is low THEN RelationQuality is regular
if CoauthorsAverage is low AND NumberOfPapers is low AND RelationTime is high THEN RelationQuality is low
if CoauthorsAverage is low AND NumberOfPapers is high THEN RelationQuality is high
if CoauthorsAverage is high AND NumberOfPapers is low THEN RelationQuality is low
if CoauthorsAverage is high AND NumberOfPapers is high THEN RelationQuality is regular

RelationTime is important in case of low coauthors average and low number of papers. In this situation, RelationTime is used to determine if quality is low or regular.

Figure 3 shows an example according to the fuzzy model. In this example, u_4 and u_5 have a relation quality of 4.68. This value is the output of the fuzzy model considering the input variables and rules presented.

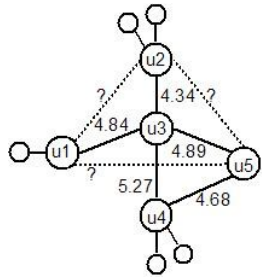


Figure 3. Relation quality using fuzzy model (Fuzzy1)

Using metric (Fuzzy2)

The second approach uses the following metric to determine the RelationQuality. The value a paper adds to the total number of papers is obtained as follows:

$$\text{PaperYear} - \text{InitialYearOfTraining} \quad (8)$$

$$C_o = \frac{\Gamma(A) + \Gamma(B)}{2} - (\Gamma(A) \cap \Gamma(B)) \quad (9)$$

$$\text{RelationQuality} = \frac{p(A, B)}{C_o} \quad (10)$$

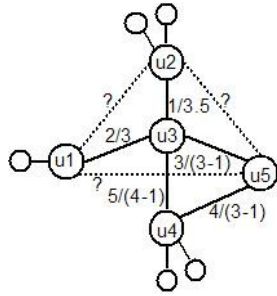


Figure 4. Relation quality using metric (Fuzzy2)

$p(A, B)$ is the number of papers co-authored by A and B , considering different weights according to the paper's year. Recent papers are more important. Figure 4 shows an example according to the used metric, where users u_4 and u_5 have value 4 to co-authored papers. This value indicates that they produced 1 paper in the last year of training in an interval of 5 years. Value 3 is the average of co-authors of users u_4 and u_5 and 1 is the number of common co-authors.

The next section presents the approach proposed to analyse the results of Link Prediction methods. This step is under construction.

3.3 ROC Curve

According to Prati [19], the ROC analysis is a graphical method to evaluate diagnostic and prediction systems. The ROC graphs were initially proposed to analyze the quality of signal transmission [6]. Nowadays they are used as a powerful tool to evaluate classifiers in Machine Learning and Data Mining areas [3], [21].

The ROC curve is obtained from the rate of false positives and true positives. This way, it is possible to compare these values in various cutoff points, not just considering a single threshold. A measure often used to evaluate classifiers is the Area Under the Curve, which can range from 0 to 1, and the greater the value the better is their performance.

In the context of Link Prediction the main goal is to determine the existence of a link between two entities. In order to do so, the link weight is assigned to each pair of nodes x and y , and then a list is generated in decreasing order of *score*. This value can represent the membership degree of the link to the fuzzy set Positive. Thus, given that the methods provide a value representing the weight of the link and not only its existence, you can take a fuzzy approach to generate the ROC curve and evaluate the Link Prediction methods.

The methodology adopted in the experiment to evaluate the process defined is presented in the next section.

4 EXPERIMENT METHODOLOGY

The adopted methodology is similar to the one presented in Liben-Nowell [14]. Suppose we have a social net $G = (V, E)$ in which edge $e = (u, v) \in E$ represents the total of interactions (co-authored papers) between u and v at different times. Having times t_0, t'_0, t_1, t'_1 and assuming that $t_0 < t'_0 < t_1 < t'_1$. $[t_0, t'_0]$ the training interval and $[t_1, t'_1]$ the test interval are considered. Let $G[t, t']$ consist of all edges in t and t' . Thus, an algorithm has access to the network $G[t_0, t'_0]$, generates a list of links that are not present in $G[t_0, t'_0]$ which needs to be verified in $G[t_1, t'_1]$.

Liben-Nowell [14] observed that the evaluation of Link Prediction methods use parameters k_{training} and k_{test} and assumed that the Core set are nodes that belong to at least k_{training} links in $G[t_0, t'_0]$ and at least k_{test} links in $G[t_1, t'_1]$. In our work, we consider Core set as nodes selected by the use of Fuzzy Sensors which consider the variables NumberOfPapers and NumberOfCoauthors.

The training interval is denoted as $G = (A, E_{\text{old}})$ and E_{new} is used to denote the link set (u, v) , u and $v \in A$. Let u and v co-author a paper during the test interval but not in the training interval ($E_{\text{new}} = A \times A - E_{\text{old}}$). These are the new interactions sought to be predicted.

Each link predictor p produces a list L_p of pairs $A \times A - E_{\text{old}}$ in decreasing order of score. For the evaluation, we focus on the Core set, thus we denote $E_{\text{new}}^* = E_{\text{new}} \cap \text{Core} \times \text{Core}$ and $n := |E_{\text{new}}^*|$. The performance measure for predictor p is determined as follows:

Considering the first n pairs in the list L_p in $\text{Core} \times \text{Core}$ and determining the size of the intersection of this set of pairs with the E_{new}^* set.

5 DATASET

The experiments were performed according to the methodology presented in the previous section. DBLP (Digital Bibliography & Library Project) is the dataset used in the experiment. This dataset contains data of Computer Science publications and has been used in different works [9], [23].

The DBLP Computer Science Bibliography from the University of Trier contains more than 1.15 million records. The DBLP contains details from publications of conferences proceedings related to Data Mining, Databases, Machine Learning and other areas. The dataset is public and is in XML format [5].

6 RESULTS AND EVALUATION

The methods based on proximity measures and the fuzzy approaches proposed in this paper must be evaluated by comparing their performance measure.

Table 2. Training and Test Periods

Period	Training	Test	$ E_{old} $	$ E_{new} $	$ E_{new}^* $
1	1999-2004	2005-2007	663530	745629	52145
2	2000-2006	2007-2009	1057817	523844	42394

Table 3. Number of Authors and Papers in Periods

Period	Authors in Training	Papers in Training	Authors in Test	Papers in Test
1	303615	377143	330324	338712
2	416808	551713	277499	245252

Table 2 shows the periods used in the evaluation and the information about dataset in periods. E_{new}^* represents the link number obtained by the use of Fuzzy Sensors. Table 3 presents the additional information about dataset in periods. Two periods were used to evaluate the methods considering different numbers of authors and publications. Period 2 has a greater number of authors and publications compared to period 1.

The results obtained are presented in Table 4. Both approaches use fuzzy compositions to determine new link weights. Fuzzy1 considers the fuzzy model to determine RelationQuality and the Fuzzy2 uses a metric.

The results show that the use of fuzzy composition represents an adequate strategy to tackle the Link Prediction problem. The use of a fuzzy model to determine the RelationQuality is interesting because it allows the use of specialist knowledge and can be adequate to different domains, considering additional or other variables.

The use of the metric presented is very simple and considers that the quality of the relation between two authors is higher in the following situations:

- When two authors have a large number of papers, mainly in recent years.
- When the average of co-authors of the authors in the relation is low, but the common co-authors are not considered, as they influence the relation in a positive way.

Table 4. Results obtained in different periods

Period	Common Neighbors	Preferential Attachment	Adamic/Adar	Fuzzy1	Fuzzy2
1	3.87	0.33	4.56	5.83	5.68
2	3.83	0.16	4.58	5.71	5.58

Table 4 shows that the Preferential Attachment method presents worse performance in both periods. According to Murata [16], the method is not appropriate for networks which have degree distributions that are almost uniform. Adamic/Adar performs better than Common Neighbor in both periods. The metric used in Fuzzy2 performs better than Fuzzy1 in both periods. The Fuzzy approach (Fuzzy1) used to obtain the RelationQuality is very intuitive and has presented better results than traditional methods.

7 CONCLUSION

The results show that the application of a process based on non dichotomic metrics in order to evaluate the methods of Link Prediction makes it possible to use the specialist’s knowledge and to adopt a perspective more similar to the human perception of the problem.

The use of Fuzzy Sensors in data selection step is very simple and intuitive and can be adequate to different domains considering different variables. Fuzzy compositions represent a satisfactory strategy for the problem of Prediction Links. The use of a fuzzy model to determine the RelationQuality is interesting because it allows the specialist knowledge in the field to be exploited in the definition of variables, since some features are peculiar to the type of social network. ROC curve can be used to analyse the results, but this step is under construction.

REFERENCES

- [1] Adamic, L. A., Adar, E., "Friends and Neighbors on the Web," Social Networks, 25(3), pp. 211-230, 2003.
- [2] Barabasi, A. L.; Jeong, H.; Eda, Z. N.; Ravasz, E.; Schubert, A.; Vicsek, T. Evolution of the social network of scientific collaboration. Physica A, v. 311(3-4), p. 590-614, 2002.
- [3] Bradley, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognition, v. 30, n. 7, p. 1145-1159, 1997.
- [4] Davidsen, J., Holger, E., Bornholdt, S., "Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks," Physical Review Letters, 88(12), p. 128701, 2002.
- [5] <http://dblp.uni-trier.de/xml/>
- [6] Egan, J. P. Signal detection theory and ROC analysis. New York, USA: Academic Press, 1975.
- [7] Getoor, L., Diehl, C. P., "Link Mining: A Survey," SIGKDD Explorations Special Issue on Link Mining, 7(2), pp. 3-12, 2005.
- [8] Glen, J., "Simrank: a measure of structural-context similarity" Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 538-543, 2002.
- [9] Hasan, M. A., Chaoji, V., Salem, S., Zaki, M., "Link Prediction using Supervised Learning," Workshop on Link Analysis, Counterterrorism and Security (with SIAM Data Mining Conference), 2006.
- [10] Huang, Z., "Link Prediction Based on Graph Topology: The Predictive Value of the Generalized Clustering Coefficient," Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (LinkKDD2006), 2006.
- [11] Jensen, D., "Statistical Challenges to Inductive Inference in Linked Data," Seventh International Workshop on Artificial Intelligence and Statistics, 1999.
- [12] Jin, E. M., Girvan, M., Newman, M. E., "The structure of growing social networks," Physical Review E, 64(4), 2001.
- [13] Katz, L., "A new status index derived from sociometric analysis," Psychometrika, 18(1), pp.39-43, 1953.

[14] Liben-Nowell, D., Kleinberg, J., "The Link-Prediction Problem for Social Networks," *Journal of the American Society for Information Science and Technology*, 58(7), pp.1019-1031, 2007.

[15] Mauris, G.; Benoit, E.; Foulloy, L. Fuzzy symbolic sensors - from concept to applications. *Measurement*, v. 12, n. 4, p. 357-384, 1994. ISSN 0263-2241. Disponível em: <<http://www.sciencedirect.com/science/article/B6V42-47YRV6C-6/2/bba07d339e347b07a06c370afb0e8f21>>.

[16] Murata, T., Moriyasu, S., "Link Prediction based on Structural Properties of Online Social Networks," *New Generation Computing*, 26(3), pp.245-257, 2008.

[17] Newman, M. E., "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences USA*, 98(2), pp.404-409, 2001.

[18] Newman, M. E., "Clustering and Preferential Attachment in Growing Networks," *Physical Review Letters* E, 64 (025102), 2001.

[19] Prati, R. C., Batista, G. E. A. P. A., Monard, M. C. Curvas ROC para a Avaliação de Classificadores. *Revista IEEE América Latina*, v. 6, p. 1-8, 2008.

[20] Popescul, R., Ungar, L. H., "Statistical Relational Learning for Link Prediction", 2003.

[21] Spackman, K. A. Signal detection theory: Valuable tools for evaluating inductive learning. *Proceedings of the 6th Int. Workshop on Machine Learning*, p. 160-163, 1989.

[22] Taskar, B., Wong, M.-F., Abbeel, P., Koller, D., "Link Prediction in Relational Data," *Neural Information Processing Systems Conf.*, 2003.

[23] Wang, C., Satuluri, V., Parthasarathy, S., "Local Probabilistic Models for Link Prediction" *Data Mining, IEEE International Conference on*, pp.322-331, 2007.

[24] Zadeh, L. A., "Fuzzy Sets," *Information and Control*, 8(3), pp.338-353, 1965.

Social Aspects of Video Recording

Alessandro Basso, Marco Milanese, André Panisson¹

Abstract. Recently, Personal Video Recorders (PVRs) are gaining success among web applications. In this paper we give some preliminary results on a framework for identifying communities on the basis of the explicit usage of a PVR. Users programming their recordings are clustered and aggregated together and social interactions among users come out from the grouping of common interests, for sake of recommendation purposes. Items recorded by similars users (within a certain measure) are ordered for recommending unseen broadcasted programs. We exploit these interactions in the Facebook social network, with an application where direct communications between users contribute to our knowledge gathering. A detailed analysis of one-year observations is given, along with some considerations about the implicit user profiling, showing that it is possible to identify groups of users with similar interests, despite the lack of explicit information about their likes or dislikes.

1 Introduction and Motivation

In the last years, we have been seeing to an always increasing convergence trend between traditional media, such as the Television, and the Internet, which resulted in a huge amount of available contents to the end-user. Common examples of such a phenomena are identifiable in services such as IP-TV, VideoOnDemand or YouTube, which have undergone a noticeable diffusion due to the high bandwidth availability granted by the last generation of Internet connection technologies.

In such a context, the need of being able to access a large amount of multimedia contents, without requiring the user to be constantly in front of the media device, has enabled the emergence of a new generation of Personal Video Recorders (PVRs) [1]. Differently from a common video recorder, a PVR is basically a software package allowing users to record audio and video contents by properly setting their time intervals (e.g., MythTV², SageTV³, MediaPortal⁴). This feature, though, comes commonly without the possibility to involve the user in a more interesting and enjoyable experience, like other web applications do, for example, by implementing user personalization through a *recommender system*, and, thus, by suggesting the user contents specifically tailored to her interest [8].

However, the intrinsic high dynamism of the PVR context, where contents may be registered and viewed by users with a daily frequency, presents a series of difficulties to the direct application of common recommendations algorithms. Aware of the known unreliability of Electronic Programming Guides (EPGs), we rely only on users' ability in precisely determine the timings of each recording.

As we do not have a reliable source of information to know that there is a specific movie on a specific channel, we only know that several recordings have been set within a certain range of timings on that channel. In this perspective, we can not deal with predefined *discrete* and *permanent* objects. Instead, we use events characterized by a specific temporal validity - the users' recordings - to extract information about which events are being recorded, on a bottom-up approach to determine the discrete events.

Our goal in this paper is to provide a social dimension to the very dynamic PVR-services domain, which is typically focused on the single user. By using events discretization combined with a collaborative filtering (e.g., [8]) approach to make good predictions of user's recordings, we aim to aggregate users with similar interests with the purpose of improving their overall multimedia experience. Our approach is based on the analysis of real data generated by the Faucet PVR system⁵, integrated in a web-based podcasting service named VCast. Faucet allows users to record favorite TV and Radio (Italian) programs, that can be further downloaded into their devices (e.g., iPod, PC, notebook).

The user can set up her own programming and see or download her recordings by the use of a simple web interface. Bringing the ability to record and group into a single feed public and private channels (such as radio and TV recorded programs), Faucet PVR offers a single framework for creating and aggregating personal podcast compilations. This system produces a very rich and dynamic data set, populated by real users expressing their preferences through the recorded programs. If such a collaborative filtering approach is able to make good predictions of the recordings, it could be immediately applied in a recommender system tool for implementing user personalization. We start here from our previous analysis given in [3], in which a detailed description of the events' definition is given.

This paper is organized as follows. Section 2 gives a brief introduction on how to extract social behavioral patterns from the explicit usage of a PVR application. Next, it is explained (Section 3) how these information can be used for aggregating items and exploited (Section 4) for recommendation purposes. Section 5 presents some early results of our analysis. Finally, Section 6 traces the route for our future works in this subject.

2 Context

The goal of a recommendation system in the PVR context is to suggest a customized set of transmissions to the users [9]. However, data coming from the Faucet PVR are not immediately usable to identify permanent and discrete events (i.e., transmissions), but assume the form of unstructured information, which has to be properly processed. In particular, let T be the set of transmissions during a day and t_i be a specific transmission broadcasted on channel c_{t_i} , starting

¹ Department of Computer Science, University of Turin, Italy, email: alessandro.basso@di.unito.it, marco.milanese@di.unito.it, andre.panisson@di.unito.it

² <http://www.mythtv.org/>, Last Visited 10 Feb. 2010

³ <http://www.sage.tv.com/>, Last Visited 10 Feb. 2010

⁴ <http://www.team-mediaportal.com/>, Last Visited 10 Feb. 2010

⁵ <http://www.vcast.it/faucetpvr/>, Last Visited 10 Feb. 2010

at time b_{t_i} and ending at time e_{t_i} . Then, t_i can be directly used in the recommendation engine and this property holds $\forall t \in T$.

On the contrary, data collected by the Faucet system (i.e., users' recordings) differ from t_i in the sense that they are a set R of several different recordings r_i with a temporal validity, each referring to a specific content. However, recordings with different properties (e.g., start time, end time, title and so on) may refer to the same content: $r_i \neq r_j$ even if r_i, r_j refer to the same transmission. Clearly, this property does not hold with discrete and well defined events such as the transmissions: $\forall t \in T, t_i \neq t_j$.

Respect to other works in this domain (e.g., [6]) we have to face the periodicity of a recommendable item. This means that, for example, given a weekly broadcasted program, we can not recommend the next-week event to the user, since the programmed recording explicitly contains this information.

A noticeable property characterizing the context in which we operate is the lack of a well defined programming for recorded contents. Despite several EPG sources do exist, we can not consider them reliable enough to be used to extract the input information of the recommender engine. Thus, we opted for a different approach: exploiting the so-called *wisdom of the crowd*, we rely on users' knowledge to define: *what* contents are broadcasted on the major (Italian) TVs and radios, *where* (the channel) and *when* (the timings).

The task of defining the parameters related to every recorded transmission, such as channel, title, timings and periodicity, is therefore demanded to single users. Since it is their interest to make sure that the information inserted in the Faucet PVR are as much precise as possible, we can consider such data reliable enough to be used in the recommendation process. As result of this users activity, we can obtain a series of recordings, defined as follows:

$$r_i = \langle u_{r_i}, c_{r_i}, tl_{r_i}, b_{r_i}, e_{r_i}, p_{r_i} \rangle$$

where u_{r_i} is the user who set the recording, c_{r_i} is the channel, tl_{r_i} is the title, b_{r_i} and e_{r_i} are the the starting and ending time respectively, and p_{r_i} is the periodicity.

One of the main tasks of the proposed framework is the *discovery* of events from the amount of unstructured data obtained through the recordings. A *discrete event* can be defined as the representative of a broadcasted transmission, obtained as the result of the aggregation of several different recordings. More formally, given the set R of recordings $r_1, \dots, r_n$, we cluster every recording possessing the same attributes for channel and periodicity but similar timings, i.e., $\forall r_i, r_j \in R$ so that $c_{r_i} = c_{r_j} \wedge p_{r_i} = p_{r_j} \wedge |b_{r_i} - b_{r_j}| > d_b \wedge |e_{r_i} - e_{r_j}| > d_e$, where d_b, d_e are the clustering diameters for the starting and end times respectively. The result is a series of clusters, whose centroid is chosen as representative. Finally, through the aggregations of similar clusters, we can obtain a number of discrete elements, defined as follows:

$$d_i = \langle \{u_{d_i}\}, c_{d_i}, tl_{d_i}, b_{d_i}, e_{d_i}, p_{d_i} \rangle$$

where $\{u_{d_i}\}$ is the list of users who set a recording referring to that transmission, c_{d_i} is the channel, tl_{d_i} is the title chosen among those given by users, b_{d_i} and e_{d_i} are the the starting and ending times respectively, and p_{d_i} is the periodicity. For more details, please refer to [3].

An example of the highly dynamic environment of PVRs is given in Figure 1, where we show all the starting times referring to those recordings which contribute to define a single discrete event (i.e., a transmission).

It is important to underline that no information on the content of recordings is used to define a discrete event. Indeed, the Faucet PVR

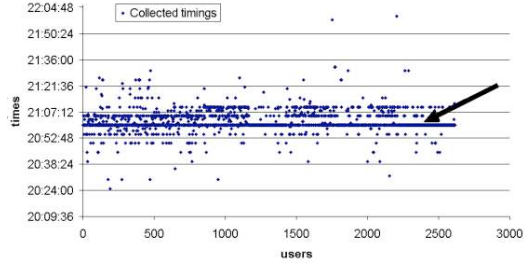


Figure 1. More than 2600 users programmed the recording of this specific transmission. As it turns out, several start times were given. The arrow indicates the representative timing we choose.

does not provide any reference to the type of transmission recorded (e.g., sport program, news, or comedy-movie), nor we can rely on information inserted by users in titles, being characterized by a high variability. Thus, we can only discriminate between two programs (on the same channel) by clustering on top of timings, plus the periodicity of the recording. In addition, the lack of information about the content of the recordings does not allow us to apply content based recommenders, which focus on the description of item and user profiles in order to recommend items [7]. On the contrary, our approach only relies on the behavior and characteristics of large interconnected networks, by exploiting similarities between their users.

3 Aggregating recordings

As told in the previous section, each recording is characterized by a *channel*, a *timing*, a *periodicity* and a *title*, a completely uncontrolled and free user generated tag. Recordings are grouped together by channel and clustered on the timing: if two recordings on the same channel present similar start and end times, then we cluster them together. Neither the size of the cluster, nor the final number of clusters is fixed *a-priori*. At the end of the clustering process, the recordings minimizing the timing differences within the computed clusters are taken as centroids of the clusters.

The aim of the clustering phase is the transformation of the recordings domain to a domain of *discrete events*: the set of individual recordings must be transformed into a set of distinct *broadcasted programs*, i.e., from the continuous domain of timings to the discrete domain of programs. Still the clustering itself is not enough. After the clustering, attributes of the discrete events are computed from the centroid of each cluster. Specifically, timings are computed by considering the median of the timings extracted from every element of the cluster, whilst the most frequent title in the cluster is chosen as a representative of that discrete element.

The aim of the aggregation procedure is to correct possible inaccuracies in the clustering phase. In fact, as in old VCR programming, most of recordings are also loosely timed in the PVR domain, meaning that the user usually started the recording a little bit earlier to stop it a little bit after, to ensure the completeness of the desired TV show. For instance, the reader may think of a TV program starting at 8 pm. It is likely to observe a large number of recordings starting in the interval [7.50, 8.00], whilst others might start in the interval [7.40, 8.10]. Due to this fact, it is problematic to directly treat both underestimation and overestimation of the timings in the clustering phase, as two possible cases can occur:

1. by choosing a larger clustering diameter (e.g., 60 minutes), we allow a more complete aggregation of recordings referring to the same content, with the side effect of possibly *skipping* the discretization of very short transmissions (e.g., whose duration is around 15 minutes).
2. By defining a shorter clustering diameter that barely takes into account underestimation and overestimation (e.g., 10 minutes), we end up with a larger number of clusters, which are likely to generate new discrete elements rather than being merged with existing ones.

In the current implementation of the aggregation algorithm, we opted for the first case, due to the low probability of occurrence of very short transmissions. The final creation of a discrete element is obviously constrained by the possible past insertion of a very similar discrete event, from which the parameters of the currently computed discrete event are updated.

Two main problems were faced in the clustering and aggregation procedure. The first, as said before, is the fact that the end user has an almost complete control on the system, thus she can set any timing and any title she wants. The channel is the only field that is given. This leads to a wide variety of titles and timings for recording the same program, but, while for the timings this was solved with the introduction of various thresholds, for the string representing the title it was necessary to compute the most frequent title of the cluster. The second problem is the periodicity of the recordings. As long as five options are available in VCast (i.e., daily, weekly, mon-fri, mon-sat, no-repeat), it was necessary to develop five different aggregating functions for managing the different behavior of the recordings.

4 Recommending events to users

The first outcome of the aggregation procedure is the daily top hits chart, namely the *most popular* events, computed with respect to the periodicity of the associated recordings. We collect all the recordings, cluster them together into discrete events and order them on the number of distinct users aggregated on the discrete events; that is, for each periodicity, we compute the most programmed discrete events, and we propose them to the users. The introduction of the most popular events appreciably increased the total number of recordings (see Figure 5), as users have no more to search on EPGs or on newspapers the broadcasted programs: most of them are included in the given chart.

The collaborative filtering approach used in grouping users can be exploited to build a valuable dataset for the setting up of a recommendation engine [2]. That is, as users are aggregated together into groups with similar interests (i.e., users have recorded the same events) then the basic idea on which layering the recommendation is: what programs have been recorded and programmed in the future by users who share some interests?

In order to define such a similarity notion between users, we introduce a *similarity function* $S(u, v)$, that returns a similarity evaluation for the given couple of users (u, v) , estimated on top of past recording activities (i.e., resources in common), programmed recordings (in the future) and similar recording behaviors. The *similarity network* among users is represented by a graph where nodes are users and there exists a weighted edge between u and v iff $S(u, v) > 0$.

Remember that we do not have any information about the proposed content, but the collaborative filtering approach allows us to group users on the basis of their personal interests, even without this knowledge. Of course this process can be improved with a feedback

mechanism that allows users to rate the suggested programs, specifying if it was satisfactory or not.

4.1 Social Networking

We exploit our framework in the Facebook social network, with the development of a VCast application⁶. This application acts both as a front-end to the PVR and as the presentation layer for our recommendation algorithm (see Figure 2) [4].



Figure 2. Recommendation front-end in Facebook.

Moreover, exploiting the inner relationships built on the social network, we can make use of a more *personalized* recommendation service. We can thus distinguish from two distinct recommendation approaches, both followed by our framework. It is important to underline that (possibly) there is no relation between the friendship on the social network and the similarity in the VCast's users network.

Direct recommendation is made on top of the friendship on Facebook and is totally unbound from the similarity in the VCast network. Users who were previously *friends* on Facebook and share the VCast application, can openly recommend each other TV shows they like.

On the other hand, *indirect recommendation* is built on top of the VCast similarity network. User's interests and past recordings are collected and computed, in order to build a neighborhood set with users with similar interests. From that set, we can then extract shows that should be of any interest for the user.

5 Results

What we have reached so far can be summarized as follows:

- Search & discovery process;
- VCast community investigation by means of network analysis;
- Recommendations based both on the community and on user's interests starting from the output of the discretization process.

VCast's users activity is studied through the number of programmed recordings, the number of downloaded files and the number of active users. Figure 3 shows the activity on a system in a year over a weekly basis. Each point in the graph represents the number of recordings (square-dotted line) and the number of downloaded events

⁶ <http://apps.facebook.com/vcastapp/>, Last Visited 19 Jan 2010.

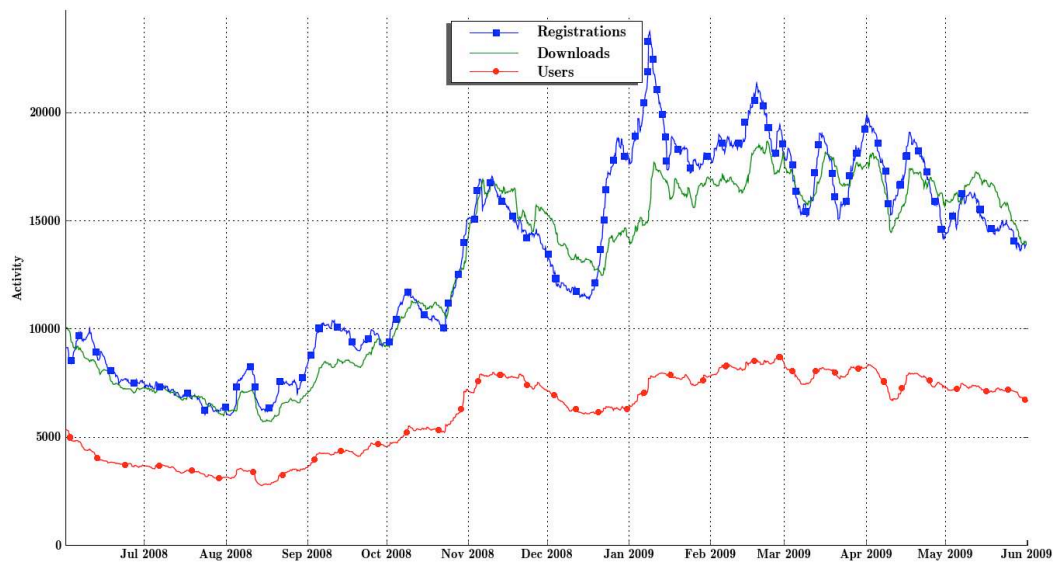


Figure 3. The activity of the system in the period: June 2008 - June 2009, expressed by means of the number of registrations, downloads and users.

occurred in the system in a 1-week interval. Just the same, each point of the circle-dotted line represents the number of active users on a 1-week interval.

As it turns out, this activity follows a constant behavior, alternating phases of increasing and decreasing programmed recordings (i.e., the peaks in Figure 3). Quite obviously, after a local maximum peak in recordings, follows a local maximum peak in the downloading activity. It is also interesting to notice that the curves decrease in the summer period, according to the reduced activity during the vacation time.

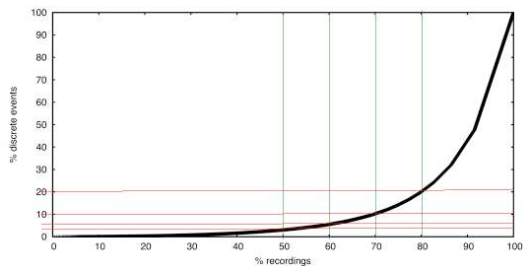


Figure 4. The long-tail effect (from 01/24/2008 to 05/18/2009: we identified 111.441 distinct events from 682.019 distinct recordings.)

As we expected, the relationship between the recordings and the identified discrete elements follows the so-called “long-tail” behavior (see Figure 4). Most recordings (about 80%) are relative to a

well defined subset of the overall events (20%).

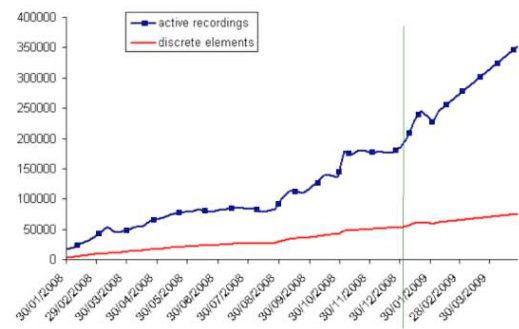


Figure 5. Since the most populars were showed on the web site, there is a significant increase in the number of distinct recordings.

This behavior is much more noticeable considering the timing in which we started showing the *most popular* recordings on the VCast web site. As Figure 5 shows, the number of recordings increased since the most popular introduction: it is an easy way to get all the relevant information on the most successful broadcasted shows, i.e., the most popular events.

The set of discrete events obtained by the discretization process is used as input for two different algorithms, devised to recommend

contents to the users of the system. In particular, the set of the *most frequently recorded* events is presented to the users as a community-based recommendation, by means of the *Most Popular* algorithm.

Similarly, a further recommendation scheme can also be devised from the discrete events, with the purpose of suggesting contents specifically tailored to users' interests. The algorithm, namely *Recommendation for Recordings* (Rec^2), exploits the well-known *collaborative filtering* approach of the *nearest neighbors* by identifying the k users $v_1, \dots, v_n$ who can be considered most similar (with respect to a chosen similarity metric) to a specific user u , i.e., neighbors of u . From the elements of interest of such users, we can then identify those which can potentially be of any interest for u , thus recommendable to her [5].

In our tests, we chose the *Dice's similarity coefficient* to measure the similarity among the sets of items E_u, E_v watched by users u and v . Other similarity measures were also tried, however the obtained results are comparable with the Dice's one, which is defined as follows:

$$S(u, v) = \frac{2|E_u \cap E_v|}{|E_u| + |E_v|}$$

In the kNN algorithm, the weight of an element e_i for a user u_k can be defined as:

$$w(u_k, e_i) = \sum_{u_a \in N(u_k)} r(u_a, e_i) \cdot S(u_k, u_a),$$

$$\text{where } r(u_a, e_i) = \begin{cases} 1 & \text{if } e_i \in E_{u_a} \\ 0 & \text{if } e_i \notin E_{u_a} \end{cases}$$

E_{u_a} is the set of elements recorded by user u_a , whilst $N(u_k)$ is the neighborhood of user u_k , limited by considering only the top- N neighbors ordered by user similarity. We explored different variants of the scheme and the value $N = 300$ yielded to the best results.

In order to get an evaluation of our discretization process, we compute precision and recall of the Most Popular and Rec^2 algorithms. Starting from the Most Popular algorithm, we have to consider two sets: $V(u, t)$, i.e., the set of events registered by user u from time t onward, and $MP(n, t)$, i.e., the set of the n most popular events computed by our algorithm at time t . It must be clear that the Most Popular algorithm calculates $MP(n, t)$ based only on the events registered before time t , without any knowledge about the events $V(u, t)$ registered after time t . In this scenario, precision and recall for each user $u \in U$ are computed as:

$$Precision(u, n, t) = \frac{|V(u, t) \cap MP(n, t)|}{|MP(n, t)|}$$

$$Recall(u, n, t) = \frac{|V(u, t) \cap MP(n, t)|}{|V(u, t)|}$$

The precision and recall at time t are computed as the average of all users' precision and recall values. $Precision(n, t)$ and $Recall(n, t)$ represent the precision and recall at time t using the n most popular events. Finally, we compute the system precision and recall at different times, and calculate the system overall precision and recall as the average of it. $Precision(n)$ and $Recall(n)$ represent the system overall precision and recall using the n most popular events. At each run, our algorithm computes n of the most popular events, with n varying from 1 to 30, and the associated $Precision(n)$ and $Recall(n)$.

The way we compute precision and the recall for the Rec^2 algorithm is similar to computing precision and recall with the Most Popular algorithm. The main difference is that, while most popular events are given to all users, recommendation is personalized to each user. To this aim, thus, we consider the sets $V(u, t)$ as the set of events registered by user u from time t onward, and $Rec(u, n, t)$ as the set of n recommended items to user u at time t . It's important to notice that all recommended items are *in the future*, and also that the Rec^2 algorithm calculates $Rec(u, n, t)$ based only on the events registered before time t , without any knowledge about the events $V(u, t)$ registered after time t .

$$Precision(u, n, t) = \frac{|V(u, t) \cap Rec(u, n, t)|}{|Rec(u, n, t)|}$$

$$Recall(u, n, t) = \frac{|V(u, t) \cap Rec(u, n, t)|}{|V(u, t)|}$$

The way we calculate the system overall precision and recall is the same as the Most Popular algorithm, and $Precision(n)$ and $Recall(n)$ represent the system overall precision and recall using the n recommended events.

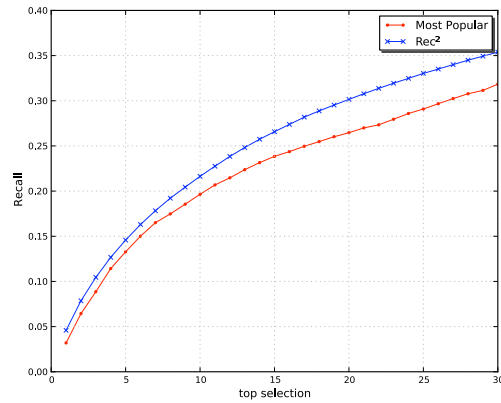


Figure 6. Recall for the recommendation algorithms.

In Figure 6 and 7, results are given. Recommending 30 events gives us more than 30% of recall with respect to the total recordings made by a single user. Regarding precision, both the recommendation algorithms score better than 11% considering the first 5 elements, which can be considered an adequate number of recommendable elements.

As it turns out, results are quite similar with the Most Popular algorithm and the Rec^2 engine, with the latter performing better in both precision and recall. This can be explained because of the lack of recommendable items due to what is depicted in Figure 4: most of users watch the same shows and nothing else. In order to retrieve previously unseen shows to be recommended we should navigate the similarity graph for farther nodes, but this lead to the widening of the set of considered items and thus increasing the risk of picking undesired content for the user.

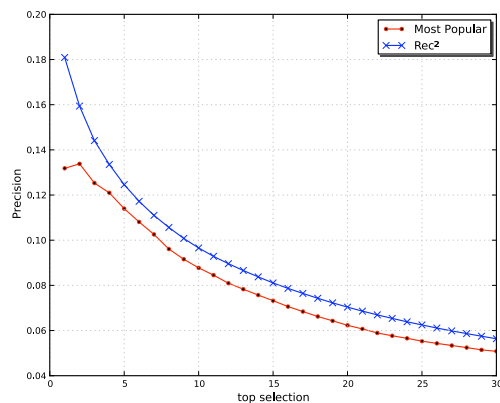


Figure 7. Precision for the recommendation algorithms.

6 Conclusion and Future Work

In this work, we present a framework for investigating the social aspects involved in the usage of a PVR application. Starting from a network analysis approach, we identify several clusters of users and sets of discrete events distinguishing the communities. Most of the users, however, tends to behave in a quite similar way, as the experience with the most popular recordings underlines.

The collaborative filtering approach gives us the opportunity to exploit an affinity function to group users with similar interests and try to predict and recommend what will be programmed for recording by the users. Still there are users who our recommendation engine is unable to recommend something to, meaning that there are no programmed recordings not already seen or programmed by that user in its neighborhood. This could lead to the identification of most hidden and isolated communities of users, unbiased from the most populars and very adherent to their *native* tendencies and interests.

Moving from the network analysis approach, it could be interesting to model this domain with a multiagent system. In such a viewpoint, users could be considered as agents which possess some *beliefs*, in the form of contents seen in the past and personal interests, and are trying to achieve specific *goals*, i.e., what they are going to program for the future. It is our intentions to investigate whether the existence of social relations among agents, modeled by means of direct (i.e., the social network) or indirect (i.e., the similarity network) interactions, could be exploited to support the agents in their activities, thus improving the overall user experience.

REFERENCES

- [1] 'Free-to-air television and other PVR challenges in europe', Tech 3314, EBU-UER, (Jan 2006). http://www.ebu.ch/CMSimages/en/tec_doc_t3314_tcm6-42262.pdf.
- [2] Gediminas Adomavicius and Alexander Tuzhilin, 'Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions', *IEEE Trans. on Knowl. and Data Eng.*, **17**(6), 734–749, (2005).
- [3] Alessandro Basso, Marco Milanesio, and Giancarlo Ruffo, 'Events discovery for personal video recorders', in *EuroITV '09: Proceedings of the seventh european conference on European interactive television conference*, pp. 171–174, New York, NY, USA, (2009). ACM.

- [4] Francesca Carmagnola, Andrea Loffredo, and Giorgio Bemardi, 'Vcast on facebook: Bridging social and similarity networks', in *HT '09: Proceedings of the Twentieth ACM Conference on Hypertext and Hypermedia*, New York, NY, USA, (July 2009). ACM.
- [5] Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers, and John Riedl, 'An algorithmic framework for performing collaborative filtering', in *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 230–237, New York, NY, USA, (1999). ACM.
- [6] Yifan Hu, Yehuda Koren, and Chris Volinsky, 'Collaborative filtering for implicit feedback datasets', in *ICDM '08: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pp. 263–272, Washington, DC, USA, (2008). IEEE Computer Society.
- [7] Michael Pazzani and Daniel Billsus, 'Content-based recommendation systems', *The Adaptive Web*, 325–341, (2007).
- [8] Paul Resnick and Hal R. Varian, 'Recommender systems - introduction to the special section.', *Communication ACM*, **40**(3), 56–58, (1997).
- [9] J. Ben Schafer, Joseph Konstan, and John Riedl, 'Recommender systems in e-commerce', in *EC '99: Proceedings of the 1st ACM conference on Electronic commerce*, pp. 158–166, New York, NY, USA, (1999). ACM Press.

Agent-based economic modeling with finite state machines

Ivan Gnilomedov (St. Petersburg State University of Information Technologies, Mechanics and Optics.
Email: zilsmail@gmail.com)

Sergey Nikolenko (Steklov Mathematical Institute, St. Petersburg. Email: sergey@logic.pdmi.ras.ru)

Abstract. We offer a novel approach to agent-based economic modeling. Previous work has modeled learning agents as neural networks, sets of fuzzy rules and other learning algorithms. In this paper, we present an approach based on representing an agent's production cycle as a finite state machine. We show that the finite state machine model offers natural representations for basic economic features such as complementary and substitute commodities. Our experimental results show that the model behaves in perfect accordance with basic economic laws that shows the model's validity.

1. INTRODUCTION

Agent-based modeling has played an increasingly important role in understanding market behaviour. Existing research has been primarily centered on modeling financial markets with agents either following predefined strategies or learning the optimal strategy via different learning algorithms.

We aim to create a general equilibrium model representing a market of companies producing different commodities with different manufacturing processes, a model that would accurately reflect the behaviour of supply, demand, and prices in a real economy. A classical model of this market is the well-known Walrasian model for N different commodities produced with M different resources (production factors), where general equilibrium is achieved by maximizing each agent's profit; see Black [1] for a detailed exposition. The general equilibrium model leads to nice mathematical properties (Walras' Law and others). However, this model assumes each agent is perfectly rational and computationally unbounded and the general equilibrium is static, while the real world market tries to achieve equilibrium in a constantly changing world. Thus, the need arises to offer agent-based models with limited and learning agents.

Zimmermann, Neuneier, and Grothmann [2] offer an agent-based model of the *FX-Market* where agents base their decisions on incomplete information coming from a limited number of error-prone sources of information. The agents are modeled as error-correcting neural networks. Kooths [3] and Kooths, Mitze, and Ringhut [4] create a macroeconomic model where agents are modeled as neural networks with additional fuzzy rules representing knowledge about the economy. Other approaches have also been tested; for a detailed survey of agent-based economy models see LeBaron, [5], Tesfatsion [6] and references therein.

Agent-based models with agents represented as finite state machines (FSMs) are widely used in computer animation (see, for example, Rudomin, Millan, and Hernandez [7]), while learning FSMs has been already applied to automata-based programming developed by Shalyto and Tükel [8].

Developing systems of reactive agents with finite automata was suggested in Naumov, Shalyto [9]. The authors assert that their approach can allow to create systems of self-learning adaptive agents.

However, as far as we know this is the first agent-based economy simulation where agents would be represented and trained as FSMs. In this paper, we fill the gap by constructing

an economic model with adaptive agents based on finite state machines. The underlying FSM of an agent represents its production cycle. It gives us an opportunity to model a production process with more detail than previous works do and to naturally interpret such relations between commodities as substitutes and complementary goods. Moreover, finite state machines are well-studied mathematical objects with many known properties, which allows us to approach the problem of devising (or learning) a suitable program for the agents.

We have implemented a model economy with these agents. Results of our experiments are in good accordance with basic economic laws. We have attempted to make the strategy that agents use to make economic decisions both easy to implement but efficient in achieving the proper suboptimal behaviour of the agents. The agents' programming (artificial intelligence) is based on the ant colony optimization algorithm which will be described below.

The paper is organized as follows. Section 2 describes the model itself: in 2.1, we show how to model a company with finite state machines, 2.2 describes the market model in our approach and shows a concrete example of a production cycle, 2.3 deals with the agents' intelligence, and 2.4 outlines the benefits of the proposed model. Section 3 lists experimental results; we show that the model behaves just as the economic laws predict, thus establishing the model's validity.

2. DESCRIPTION OF THE MODEL

2.1. Modeling production with finite state machines

A *finite state machine* is a directed graph with captions on edges $FSM = \langle V, E, \Sigma \rangle$ where V is the set of vertices, Σ is the FSM's *alphabet* (the set of input events), and $E = \{ e = \langle v_1, v_2, \sigma \rangle \mid v_1, v_2 \in V, \sigma \in \Sigma \}$ is the set of

edges (possible *transitions*). An edge $e = \langle v_1, v_2, \sigma \rangle$ means that transition from v_1 to v_2 with input event σ is possible. A finite state machine has two kinds of special vertices: a unique $s \in V$ is labeled as the *initial vertex* and some $\{ t_0, \dots, t_K \} \subset V$ represent the set of *terminal* vertices.

Sometimes it is reasonable to associate both income and outcome events with each edge. Then $Tr = \langle V, E, \Sigma, \Pi \rangle$,

$$E = \{ e = \langle v_1, v_2, \sigma, \pi \rangle \mid v_1, v_2 \in V, \sigma \in \Sigma, \pi \in \Pi \}$$

(these FSMs are usually called *transducers*). A *path* in a FSM is an ordered set

$$P = \langle v_0, e_0, v_1, e_1, \dots, v_L \rangle \mid e_0 = \langle v_0, v_1, \sigma_0, \pi_0 \rangle,$$

$$e_l = \langle v_l, v_{l+1}, \sigma_{l+1}, \pi_{l+1} \rangle, \dots, e_{L-1} = \langle v_{L-1}, v_L, \sigma_L, \pi_L \rangle$$

We call the word $I\text{Label}(P) = \sigma_0, \sigma_1, \dots, \sigma_{L-1}$ the *input label* of the path P , and by the *output label* of the path P we mean the word $O\text{Label}(P) = \pi_0, \pi_1, \dots, \pi_{L-1}$. A *cycle* is a path with identical start and finish: $v_0 = v_L$. We describe a complete working example of a production cycle FSM below; for more information on FSMs we refer to Lothaire [10].

We model the production cycle of each agent as a finite state machine. The state of the FSM corresponds to a certain stage of the production process. The initial state corresponds to the zero-stage of the production; the product then travels from one production department to another, as the FSM travels from one state to another.

Each state transition corresponds to performing a certain production stage and, as such, requires a certain amount of resources. We add an internal resource pool for each agent and consider the resources necessary to complete a certain production stage as the input event for the corresponding transition. The transition may take place only if enough resources are available.

A product travels from one production stage to another until it is ready. A complete product corresponds to a terminal state of the FSM. Then the product goes to a “warehouse” where it awaits deployment to the market. In the FSM terms, we model a certain production stage as a transition from a terminal state to the initial state; to this transition we associate the resources necessary to store and transport the complete product to the market.

Each production stage produces something; thus, we add to each edge a set of *output* resources that get produced during this transition. It is natural for finite state machines to have both input and output (entry and exit) actions; in our model, the entry action consumes resources, while the exit action produces new resources.

Besides a natural representation of a real-life company, the FSM formalism has several formal advantages. First, it easily allows an agent to be flexible. In the real world, an agent can reorganize his or her company to produce different commodities (at the very least, different kinds of a commodity). In the FSM model, we allow the FSM to be non-linear (to have forks). Depending on the path an agent takes, different commodities will be consumed and produced.

Second, while in the real world agents may adapt their strategies to changing market conditions, it takes time and resources to complete the adaptation. In the FSM model, this “transition cost” is modeled naturally: if a FSM is currently in a non-terminal state, it will take time and resources to reach the terminal state before the agent can choose a different path starting from the initial state.

Finally, in the real world the agents' behaviour is only suboptimal: agents do not possess complete information about the market and sometimes cannot compute the absolute best strategy. With this in mind, we model the agents as having rather simple strategies. An agent is modeled as a pair $\langle \text{FSM}, S \rangle$ consisting of a FSM and a strategy S .

2.2. Modeling the market

In this subsection, we describe how the market itself and interaction between agents are represented in our model. Note that [5] outlines two practical approaches to determining prices determination: *true order book* (where agents post offers to buy and sell goods) and agents' *random bump mechanism* (bumped agents trade if it benefits them both). The 29 technique proposed in this paper combines the two approaches introduced in [5]. We model the market as a virtual “bulletin

board” where each agent may post an offer indicating that he is willing to sell a certain amount of a certain product for a certain price. When an agent needs to buy a certain product, he queries the market for the selling offers on this particular product. To model insufficient information, we have the market to return only a certain random subset of the offers (the offers that this agent “knows of”). The agent then reviews these offers, chooses the most profitable, and satisfies them (buys the necessary product).

In economics, two goods are said to be *substitutes* if they can be used instead of each other. An increase of demand for the first substitute implies a reduction of demand for the second. On the other hand, *complementaries* are goods that are frequently consumed with each other, and raising the demand for the first one would also raise the demand for the second. The *cross-elasticity*:

$$E_c = \frac{Q_1^B - Q_0^B}{Q_1^B + Q_0^B} \frac{P_1^A + P_0^A}{P_1^A - P_0^A}$$

helps to describe substitutes and complementaries mathematically. Here Q_0^B, Q_1^B are demand values for good

B before and after the price of commodity A changed, P_1^A, P_0^A are prices on the commodity A . The cross-elasticity is positive for complementary goods and negative for substitutes. In this model, each agent is both a customer and a producer (people are considered as suppliers of labor), and complementary and substitute goods can be interpreted with FSMs. In the model, complementary goods will often appear on consecutive edges of the production FSM, while substitute goods will appear on edges of parallel paths of the FSM. This results in that whenever an agent buys a certain commodity, his demand for the complementaries increases (he is halfway through the production, so he needs to complete it), while his demand for the substitutes decreases (he has already come through a certain path, and he will not return to a parallel edge in this cycle).

Finally, let us present a complete example of a production cycle in our model. The FSM belonging to an agent D is depicted on Fig. 1. The initial state is marked as 0, while 3 and 5 are terminal states. Labels with $-$ and $+$ denote resource costs and resource outcome, respectively; the example has four different resources, from $R1$ to $R4$. Here is a sample production cycle.

1. *Registering offers.* Suppose that other agents $F_1 \dots F_N$

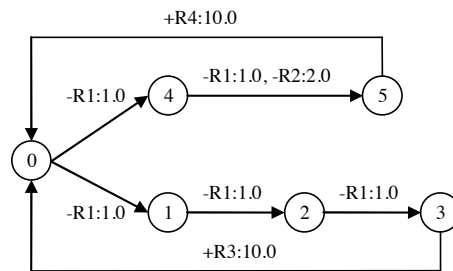


Fig 1. A sample production FSM

have registered the following offers:

```

<R1, price=1.0, amount=12.0>,
<R1, price=2.0, amount=34.0>,
<R1, price=3.0, amount=45.0>,
<R2, price=10.0, amount=67.0>,
<R3, price=4.0, amount=89.0>.

```

2. *Deciding upon the production path.* Agent *D* analyses situation on the market and decides that the path $\{0, 1, 2, 3\}$ is expected to be better than the path $\{0, 4, 5\}$.

3. *Buying resources.* To begin his production cycle, *A* needs to buy 1.0 of the resource R1. He sends a request for purchasing R1. The market randomly selects a subset of registered propositions, say,

$\langle R1, \text{price}=1.0, \text{amount}=12.0 \rangle,$
 $\langle R1, \text{price}=3.0, \text{amount}=45.0 \rangle,$

and shows this set to *D*. Agent *D* chooses the most profitable proposition, with the price of 1.0, and makes his purchase.

4. *Transition.* Agent *D* now has the necessary resources to complete the first step of his strategy and move to state 1. Note that after these steps agent *D* is completely committed to his chosen production path; he cannot switch to producing R4 before completing this round of R3, even if in state 2 he suddenly discovers that the market is desperately lacking R4, and the prices could be exorbitant. We now skip a few steps as *D* gathers all necessary resources and completes his production cycle. Suppose that *A* has just made the transition from state 3 back to state 0 and obtained 10.0 units of R3.

5. *Registering a selling offer.* As *D* now has a surplus of R3, he will be willing to register an offer to sell R3. Again, we do not precisely specify the decision rule, but as input *D* should consider both his production cost and the market situation which he is able to peek by viewing the (random subset of) offers on R3.

After this step, the production cycle repeats. *D* is now able to choose another path.

2.3. Strategies of the agents

As we have already noted, agents use an ant colony optimization (ACO) algorithm choose what and how to produce (what path of the FSM to choose). ACO is a kind of swarm intelligence algorithms based simulating an ants colony behaviour; see Dorigo and Stützle [11] for more details.

Adapting ACO for use with FSM that models production process, we obtain the following algorithm. Each agent owns a number of ants. Each ant emulates its owner's random behaviour. It means that an ant has the same production FSM as its owner, but it makes random decisions when choosing a path. If an ant gets profit, it puts pheromones on his last path, the amount of pheromone being proportional to the profit. As a result, more profitable paths contain more pheromones, and in the end the only thing the real agent has to do is to choose a path with maximum amount of pheromones. We also make the pheromones to evaporate with time in order for the agent to be able to update its behaviour if environment changes.

Here is the pseudo-code of an agent's behaviour:

```
out = maximizeFeramon(cs.getOutcomes())
market.purchase(out.getRequirements())
cs = out.getDestinationState()
for r, v in out.getProduced():
    p = choosePrice(r, market)
    market.registerSellProp(r, v, p)
```

where *cs* is a current state of agent's FSM, *market* is an implementation of virtual market described in 2.2. Implementation of method *choosePrice(...)* is guided by the formula:

$$P = \alpha \cdot Gr \cdot AC + (1 - \alpha) \cdot \max(AC, P_{alt})$$

where $Gr \geq 1$ measures an agent's greediness, P_{alt} is an expected price of the resource in the market, *AC* is the average cost of produced resource and $\alpha \in [0..1]$ measures how much the agent wishes to consider other agents' costs. We use

$\max(C, P_{alt})$ to forbid making price less than cost (being unprofitable).

2.4. Motivation of the FSM approach

Using finite state machines allows for modeling the agents' production process more accurately. It provides an opportunity to naturally model such relations between commodities as substitutes and complementary goods. It also makes agents to think a few steps forward when they make decisions.

An agent whose production is modeled with a FSM cannot switch from one production cycle to another in the middle of production because it has to wait until a terminal state is achieved. This is another real world property that gets naturally modeled in our approach.

On the other hand, FSMs are a well known and deeply studied mathematical object, and there exist many highly developed mechanisms of analyzing their behaviour. For example, the ACO algorithm is a natural fit for learning optimal strategies.

3. EXPERIMENTS

We have implemented the FSM market model and performed experiments that were aimed to check the basic economic facts about the model; they should serve as "sanity checks" for our modeling method. As a general result, the model turned out to be extremely viable, in clear-cut cases always performing just as the economic theory predicts. Thus, we expect it to have some predictive power as well, but this should be supported by further practical experiments. In this section, we describe the experiments we have carried out.

A common scenario of an experiment consists of the following steps: generating FSMs, assigning FSMs to agents, distributing initial cache among agents, introducing the *initial resources merchant* (this is an agent which sells resources necessary for all other agents to perform their first steps), running a number of steps, and collecting results data.

Generated automata look much like the one on Fig. 1 (but more complicated); they consist of a number of production cycles starting with an initial state and finishing with a terminal state. Some pairs of production cycles have a common prefix of length more than one. For example, if we add vertices 6, 7 and the cycle $\{0, 1, 2, 6, 7, 0\}$ to the automata on Fig. 1, then cycles $\{0, 1, 2, 3, 0\}$ and $\{0, 1, 2, 6, 7, 0\}$ will have a common beginning with length 3, so the agent will have to make a choice not only in the initial state. Almost all edges have nonzero income resources and zero outcome

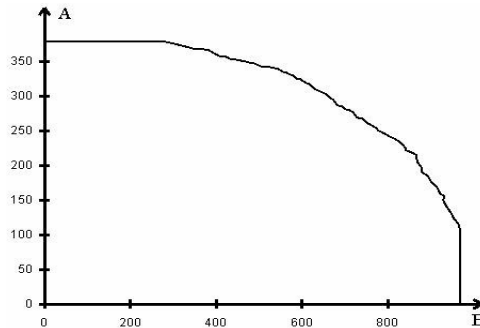


Fig 2. The production-possibility frontier

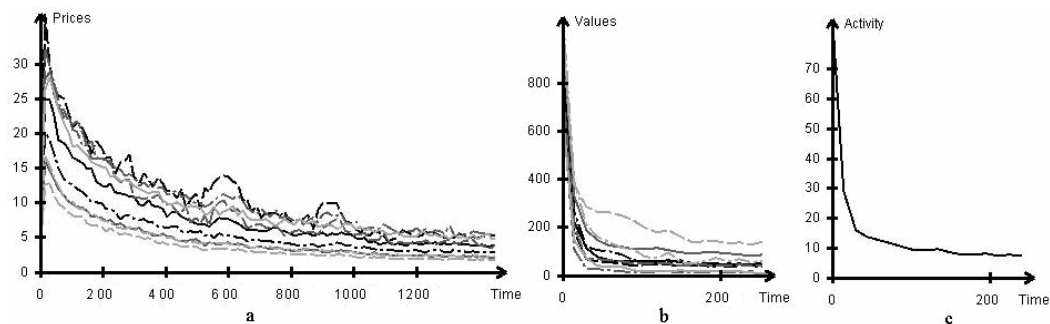


Fig 3. Overstocking: a – prices; b – gross sales; c – active agents

resources; nonzero outcome is allowed only for edges from a terminal state to the initial state. The precise way to assign income and outcome (I/O) resources depends on the purpose of an experiment.

3.1. Production-possibility frontier

The production-possibility frontier is a curve that shows the maximal volume of producing a certain commodity dependent on the level of production of another commodity. This function should be decreasing, as increasing the level of production of a certain commodity should lower the amount of resources the market is ready to assign to producing other commodities. Besides, it should be convex, as in shifting from commodity A to commodity B the market would first shift the resources that are most useful for B and least useful for A.

Our experimental scenario included two commodities: A (automobiles) and B (blankets). Each agent is able to produce both commodities, and agent's FSM has several paths corresponding to different output ratio (automobiles)/(blankets) of the commodities that this agent produces during a production cycle. Moreover, each agent has a personalized production FSM that has its own (randomly varied) production costs and A/B ratios. A certain "command center" orders some agents to switch production from A to B, and the agents (with some delay, as described above) switch production; the remaining agents act in their own interest. Fig. 2 shows the results of this experiment: the production-possibility frontier is almost convex, with small fluctuations that can be related to suboptimal behaviour of the agents.

3.2. Overstocking

In this experiment we randomly assign I/O resources to edges but the expected output of all production cycle of all agents covers expected input and has some surplus. In other words, if all agents sequentially use all their production cycles, they will not only cover all their costs of all resources, but get a profit. In the real world, after an initial surge the overproducing warehouses would be full, and prices would experience a steady monotone decrease as the overproduction continues. As prices drop, some agents should become unprofitable and should quit the market (switch to the strategy of producing and consuming nothing), which gradually compensates overproduction.

Fig. 3 depicts the results of the experiment; the graph on the left shows how prices behave with time, the graph in the middle shows how many agents are active (produce something) at this time, and the graph on the right shows the gross sales volume. Both curves behave just as predicted reach an equilibrium.

3.3. Equilibrium

In the real world, overproduction is natural, and it is naturally compensated by consumers who buy end products. To model this situation, we introduced a special agent (representing the consumers) that buys surplus resources. Fig. 4 shows that in this case, equilibrium is reached faster, and both sales and prices stabilize at higher levels.

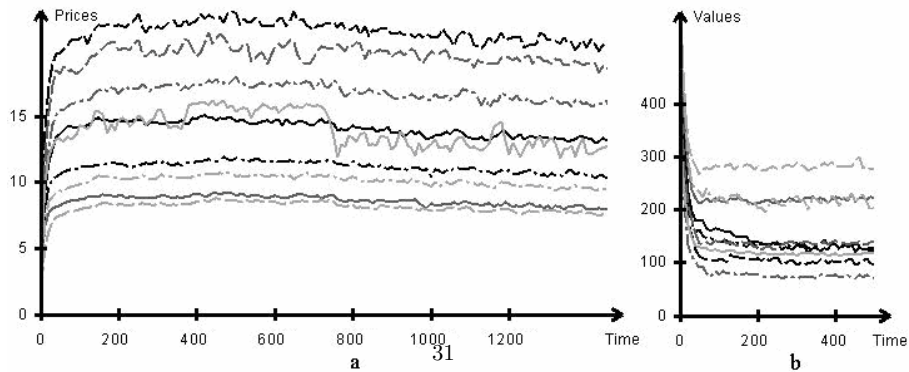


Fig. 4. Equilibrium: a – prices; b – gross sales volume

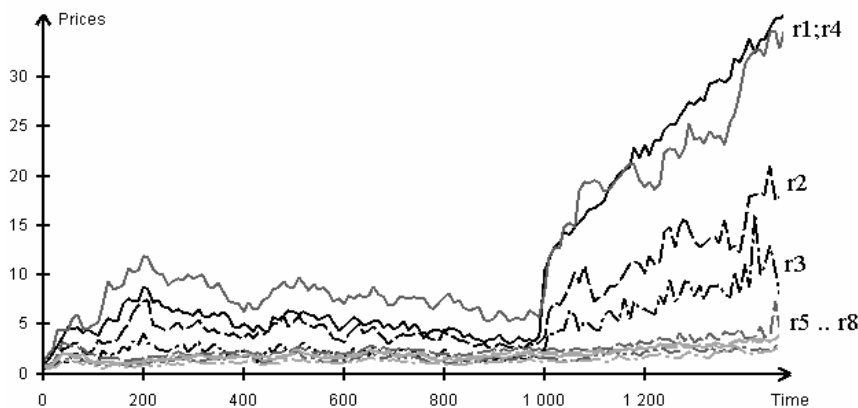


Fig. 5. Prices on a group of commodities: raw materials and end products

3.4. Raw materials and end products

In this experiment we divided nine commodities of the model into two groups: $\{r0 \dots r3\}$ and $\{r4 \dots r8\}$. We assume that production of a commodity depends strongly on other commodities of its group and to a much lesser extent on the commodities of the other group. When we generate a production cycle for an agent's FSM in this experiment, we randomly choose an output resource for the cycle. When generating input resources for a certain cycle, we choose a resource of the same group as the output resource with much greater probability than a resource from another group.

At time 1000, we introduced an agent who buys lots of $r0$, thus raising prices. As a result, prices for $r1 \dots r3$ also rise, while prices for $r4 \dots r8$ stay virtually the same. Fig. 5 shows the price graph.

3.5. Complementaries and substitutes

In the first part of this experiment, $r0$, $r1$, and $r2$ are complementaries, that is, we increase the probability that they appear on consecutive edges of the production FSMs. An increase in the price of $r0$ should cause a decrease in demand on $r1$ and $r2$.

Fig. 6 shows what happens if at time 1000 we introduce an agent who buys lots of $r0$. The market reacts by rising all prices except for $r1$ and $r2$: since producers buy less $r0$, they

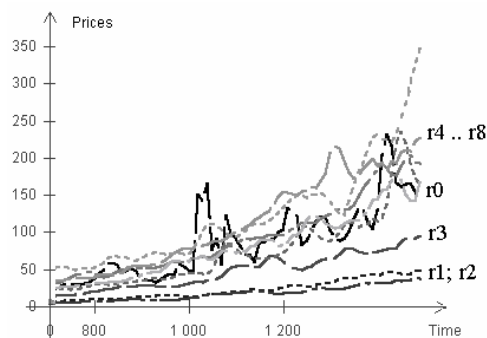


Fig. 6. Prices on complementary commodities

need less $r1$ and $r2$, too. Fig. 7a shows the same graph with only $r0 \dots r2$ visible. And Fig. 7b shows the results of the same experiment with three commodities left on the market, among which $r0$ and $r1$ are complementaries: an increase in the price of $r0$ causes an increase only in the price of $r2$.

In the second part of this experiment, we introduce $r0$, $r1$, and $r2$ as substitutes, that is, increase the probability that they appear on parallel paths in production FSMs. Here we use the following strategy for generating an agent's FSM. We begin with randomly assigning input resources for edges from the first production cycle of the FSM. When we generate an input resource for the edges from following cycles we are more likely to choose $r0 \dots r2$ if a similar edge from the first cycle has $r0 \dots r2$ as input resource. For example, if we choose $r1$ for edge $\langle 1, 2 \rangle$ from FSM from Fig. 1, then we will be more likely to choose $r0 \dots r2$ for edge $\langle 4, 5 \rangle$ than $r3 \dots r8$. Doing so, we ensure that resources $r0 \dots r2$ are alternatives, substitutes from the agents' outlook.

At time 600, we introduced an agent who buys lots of $r0$, thus increasing its price; as a result, agents begin to use less $r0$ and more $r1$ and $r2$. Fig. 8 shows the demand volumes for experiment with substitute commodities.

4. CONCLUSION

In this paper, we introduced a novel approach to modeling agents that act as producers and consumers on a market. We model an agent's production as a finite state machine with spent and produced resources as entry and exit actions. We have shown that a finite state machine is a natural way to model a company's production, and have shown how to model the basic properties of an economic agent and basic types of market commodities via finite state machines. We have implemented the proposed model and performed experiments in order to verify that the model works, and it works indeed: in all experiments the model behaved as expected.

As with any other model, we can point out the limitations of our approach. The primary limitation is that to create adequate results, the FSM agent-based model requires a large number of independent agents. This means that the model is applicable only to perfect competition markets and monopolistic competition markets.

This paper is, to a large extent, a proof of concept, evidence in support of the validity of the FSM model. Further work should deal with different learning algorithms and

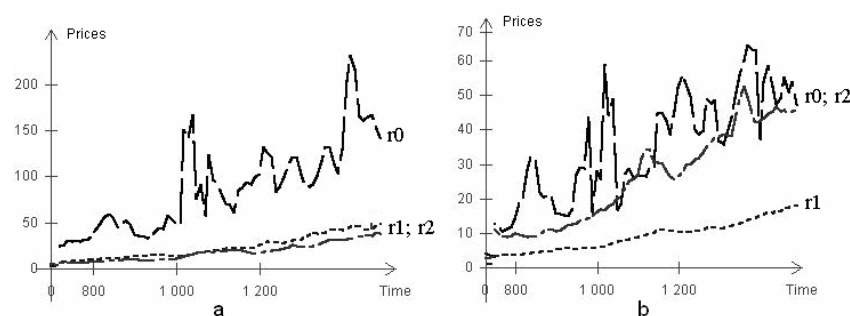


Fig. 7. Prices on complementary commodities: a – a group of three commodities; b – a group of two

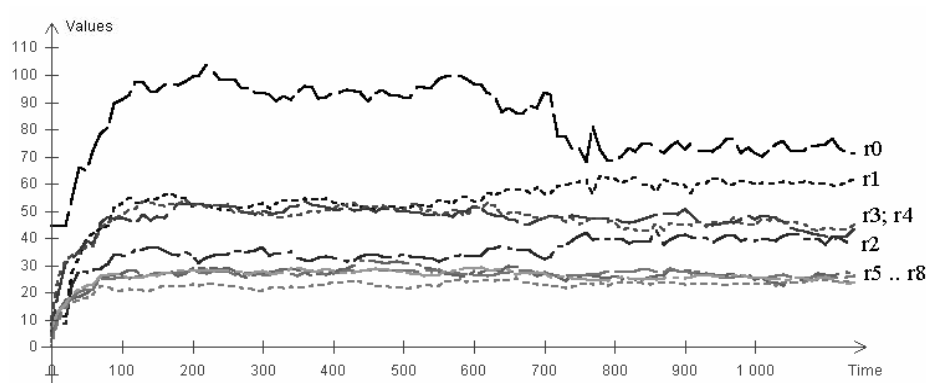


Fig. 8. Demand on substitute commodities

different pricing strategies for the agents. We plan to experiment how different strategies compete with each other and how well different FSM learning algorithms perform in this model. The model provides a natural competitive environment for testing various FSM learning algorithms against each other. And, of course, the ultimate test for our model would be to learn on real data (for example, on the stock market data) and test its predictive power.

REFERENCES

[1] F. Black. Exploring General Equilibrium. *The MIT Press, Cambridge, Massachusetts*. (1995).
[2] G. Zimmermann, R. Neuneier and R. Grothmann. Multi-agent FX-Market Modeling Based on Cognitive Systems. In: *Proceedings of the Artificial Neural Networks, ICANN 2001, Lecture Notes in Computer Science, Vol. 2130, pp. 767–774*. (2001).
[3] S. Kooths. Modelling Rule- and Experience-Based Expectations Using Neuro-Fuzzy Systems. In: *Computing in Economics and Finance – Fifth Conference of the Society for Computational Economics, Boston*. (1999).
[4] S. Kooths, T. Mitze and E. Ringhut. Forecasting the EMU Inflation Rate: Linear Econometric Versus Non-Linear Computational Models Using Genetic Neural Fuzzy Systems. *Advances in Econometrics, Vol. 19, pp. 145–173*. (2004).
[5] B. LeBaron. Agent-based Computational Finance, in *Handbook of Computational Economics, L. Tesfatsion and K. Judd (eds.)*. North-Holland, Amsterdam, pp. 1187–1232. (2006).
[6] L. Tesfatsion. Agent-Based Computational Economics: A Constructive Approach to Economic Theory, in *Handbook of Computational Economics, L. Tesfatsion and K. Judd (eds.)*. North-Holland, Amsterdam, pp. 831–880. (2006).
[7] I. Rudomin, E. Millan and Hernandez, B. Fragment shaders for agent animation using finite state machines. *Simulation Modelling Practice and Theory, Vol. 13, Issue 8, pp. 741–751*. (2005).

[8] A. Shalyto and N. Tukkel. SWITCH-Technology: An Automated Approach to Developing Software for Reactive Systems. *Programming and Computer Software, Vol. 27, No. 5, pp. 260–276*. (2001).
[9] L. Naumov and A. Shalyto. Automata Theory for Multi-Agent Systems Implementation. *Integration of Knowledge Intensive Multi-Agents Systems (KIMAS'03), IEEE Boston Section, pp. 65–70*. (2003).
[10] M. Lothaire. Applied Combinatorics on Words. *Cambridge University Press*. (2005).
[11] M. Dorigo and T. Stützle. Ant Colony Optimization. *MIT Press, Cambridge*. (2004).

The hidden path from delegation to obligations: a logical analysis

Emiliano Lorini¹

Abstract. Delegation is one of the main determinants of coordination, joint activity and collaboration within the context of organizations. In particular, delegation is fundamental for the redistribution and for the transfer of tasks within the context of an organization, and is responsible for making organizations dynamic. We present in this paper a conceptual and formal analysis of delegation. In our approach, delegation is a kind of social action that involves two agents called the delegant and the delegate, that the delegant performs by giving an order to the delegate, and that produces certain normative effects on the delegate and, in particular, an obligation for the delegate to perform the delegated task. The concept of delegation and its essential conditions will be formalized by a modal logic of agents' mental attitudes including beliefs and goals, extended by communication modalities which enable to describe what a given agent tells to another agent in front of a group of agents (an audience).

1 Introduction

When looking at social interaction within the context of organizations, social scientists have been mostly interested in individuating the antecedents of collective behavior and collective action between interacting individuals. A central concern of the field has been to identify the determinants and constituents of coordination, joint activity and collaboration. Among the different determinants, trust and delegation have been recognized as ones of the most important [2, 22, 6].

The primary function of delegation is the redistribution and the transfer of tasks, powers, and obligations, within the context of an organization. Suppose that agent i is the chief of an organization and has a general goal to achieve. Moreover i has authority over other members of the organization called agent j and agent z . Agent i might decide to exploit agent j and agent z in order to achieve his goal thereby forming a complex plan involving the actions of j and z . In this situation, i will delegate some elements in his multi-agent plan to j and z and, as an effect of the delegation act, there will be an *obligation directed* to j and z to accomplish the tasks required by i .² In this situation, i 's act of delegation to j and z is an exercise of authority by agent i over his subordinates j and z whose effect is the creation of certain obligations over j and z . Note that i 's authority over j and z (i 's right to require j and z to perform certain actions and to accomplish certain tasks) is nothing else than a specific kind of *institutional power* of agent i .

The aim of our work is to clarify the concept of delegation which is of primary importance not only for organization theory and for the design of agent organizations, but also for the design of multi-agent applications in which agents are very specialized and some of them have heavy workloads. In these situations, an agent could delegate certain tasks to other agents in order to optimize his performance and the performance of the whole system.

We provide a formal cognitive analysis of delegation by modeling interacting agents in terms of their mental attitudes (beliefs, goals, intentions) and in terms of their communicative actions. On the one hand, we define the essential *mental constituents* of delegation, that is, those mental attitudes which characterize the cognitive state of an agent delegating the performance of a certain action (or task) to another agent. On the other hand, we describe the overt social level of delegation by specifying what an agent *expresses* when he delegates the performance of a certain action (or task) to another agent.

The analysis of the delegation proposed in this work is compatible with Speech Act theory [20, 24]. Indeed, we claim that delegation requires a communicative process from the delegant to the delegate. More precisely, we claim that delegation is a kind of social action that involves two agents called the delegant and the delegate, that the delegant performs by giving an order to the delegate, and that produces certain normative effects on the delegate and, in particular, an obligation for the delegate to perform the delegated task. This relationship between delegation and obligations will be in the focus of the present analysis. We will shed light on the hidden path from delegation to obligations, by showing that agent i 's act of delegating action α to agent j necessarily gives rise to an obligation for j to do the delegated action, because a necessary precondition of i 's act of delegating action α to j is i 's power to command j to do action α .

The sequel of the article is organized as follows. In Section 2 we provide an informal analysis of delegation trying to individuate its essential constituents. This part of the work provides the conceptual backbone of all the work. In Section 3, we present a formalism which is sufficiently expressive to capture the concept of delegation discussed in Section 2. It is a modal logic of agents' mental attitudes including beliefs and goals, extended by communication modalities which enable to describe what a given agent tells to another agent in front of a group of agents (an audience). In Section 4 the modal logic of Section 3 will be used in order to define the concept of delegation, and to shed light on the logical relationship between delegation and obligations. In Section 5 we will discuss some related works on the concept of delegation.

¹ Université de Toulouse, IRIT-CNRS, France, email: lorini@irit.fr

² In the present analysis, when we talk about the concept of obligation, we just refer to the specific concept of *directed obligation* (see, e.g., [7, 14] for a general account of the concept of directed obligation): obligations which are directed from the delegate (responsible for fulfilling the obligation) to the delegant.

2 Delegation: an informal view

We conceive delegation as a specific kind of a social action involving two agents, the delegant and the delegate.

In our approach, an act of delegation is performed if and only if, the delegant orders the delegate to accomplish a certain task or to do a certain action α and the delegant has the power to command the delegate to do α . Therefore, in order to understand what *delegation* is, we have to define first the concept of *order*. The analysis of order and the related concept of request we propose is compatible with Speech Act theory.

The primary aim of Speech Act theory [20, 24] is to identify the *conditions of success* of a communicative act (or speech act)³, that is, the conditions that must be obtained in a possible context of utterance in order that the speaker successfully performs that act in that context. For example, a condition of a promise to perform a certain action in a certain context is that the speaker intends to perform this action in this context (sincerity condition). If the speaker does not intend to perform this action in this context, then he cannot make a (non-defective) promise to perform this action in this context.⁴

Let us now explain what are the different conditions that must be satisfied in order to say that a certain speaker called i has performed an order. First of all, we have to suppose that agent i has a goal to achieve and that i decides to exploit an action α of another agent j in order to achieve such a goal. In this case, agent i *relies* on action α of agent j , what is exactly described by the following Condition A:

(A) agent i wants (or intends) that agent j will perform action α .

This aspect of *reliance* is for us the first necessary constituent of directive speech acts including requests and orders, where reliance just means an agent's decision to exploit the actions of other agents in order to achieve his goals.

Rational goals and intentions must be realist [4], that is, a rational agent i cannot want something to be true unless i thinks it to be possible. Thus, Condition A implies the additional condition:

(B) agent i believes that agent j has the ability to perform action α .

Reliance is not yet sufficient to define orders and requests, although it is a fundamental dimension of directive speech acts. For instance, suppose that everyday agent i carries a pack of letters to the post office. If agent j adds a letter to the pack of letters without i seeing j , j relies on i but we cannot say that j makes a request (or gives an order) to i . Orders and requests imply a communicative dimension that it is not necessarily implied by reliance. Indeed, if an agent i orders an agent j the performance of a certain action α then, necessarily:

(C) i tells to j that Condition A is true (i.e. that he wants that j will perform action α).

³ Roughly speaking, 'speech acts' and 'communicative acts' are the same thing. Speech Act theory is indeed sufficiently general to speak about communication in general (verbal communication, communication by gestures, by facial expressions, etc.). The main objective of Speech Act theory is indeed the specification of different *types* of acts with different illocutionary forces (request, order, promise, etc.), independently from how every instance of speech act is really realized and performed.

⁴ Technically in Speech Act theory, if this condition is *expressed* through the performance of this act, this performance will be successful. Moreover if the speaker is here sincere (that is, he believes what he expresses) the performance of this act will be non defective. See [24, p. 130] for more details. Here, we do not distinguish between success conditions and non-defective conditions.

As emphasized by Vanderveken [24] sincerity conditions are necessarily *expressed* by the speaker when a speech act of directive type is performed. Conditions A and B are respectively nothing else that the sincerity condition and a preparatory condition of directive speech acts. Thus, Condition C just says that order entails the speaker's expression of the sincerity condition of a directive speech act.

It has to be noted that Conditions A and B above are *external (or overt) conditions* since they are about the open and objective level of communication (i.e. they are characterized in terms of what the speaker expresses), and are distinguished from *internal (or mental) condition* C which refers to the subjective level of communication (i.e. it is characterized in terms of the speaker's mental states). It has also to be noted that Conditions A, B and C together define the basic directive speech act *request* (see the first row in Fig. 1).

	Mental (sincerity condition)	Mental (preparatory condition)	External	Power to command
Request	A	B	C	/
Order	A	B,D	C,E	/
Delegation	A	B,D	C,E	F

Figure 1. Conditions of request, order and delegation

In order to define the concept of order, other conditions have to be added. Order has a normative dimension which is not necessarily a dimension of request. In particular, if agent i orders agent j the performance of action α then, necessarily:

(D) agent i believes that he has the power and the authority to command agent j to do action α .

Moreover, if agent i orders agent j the performance of action α then, necessarily:

(E) i tells to j that if j does not perform action α , j will violate a norm and will be liable to punishment.

Again, note that Condition E above is an *external condition* since it is characterized in terms of what the speaker expresses in the context of the communication with j . On the contrary, Condition D is a *mental condition* since it refers to the subjective level of communication (i.e. it is characterized in terms of the speaker's mental states).

Conditions A, B, C, D and E are together sufficient to define the speech act *order* (see the second row in Fig. 1).⁵

We claim here that in order to define delegation it is sufficient to add a new condition to Conditions A, B, C, D and E. Indeed, what really matters in delegation is that i has the power and the authority to command agent j to perform action α , and not simply that i believes that he has such a power. In other words, we claim here that delegation is an order by the speaker which is performed under the appropriate execution conditions (i.e. when the speaker has indeed the power and the authority to give the order). So, if agent i delegates to agent j the performance of action α then, necessarily:

(F) i has the power and the authority to command agent j to do action α .

⁵ Note that orders have also a declarative aspect: when agent i orders agent j to do action α , i declares that j has the obligation to do action α . And, if i has the power and the authority to command j to do action α then, by means of his declaration, i will create the obligation for j to do action α .

From this perspective, delegation is something more than a speech act (see the third row in Fig. 1). It is a speech act of type *order*, as defined by the Conditions A-E, which produces certain normative effects on the hearer, given the speaker's power and authority to command *j* to do a certain action (Condition F).

The following example better illustrates the six conditions A-F defining delegation.

Scenario 1 Suppose that agent *i* is the president of a company and suppose that this company must be represented at a journalist meeting. Suppose now that agent *i* delegates to agent *j*, an employee of the company, the task of participating to this meeting. In this case, the employee does have any refusal option (this is an order of the president). This means that *i* relies on *j*'s action of participating to the meeting (i.e. *i* wants that *j* will participate to the meeting) and believes that *j* is able to participate to the meeting. Agent *i* also believes that he has the power and the authority to command *j* to participate to the meeting. Furthermore, *i* openly tells to *j* that *i* wants that *j* will participate to the meeting and that if *j* does not participate to the meeting, *j* will violate a norm and will be liable to punishment. Finally, qua president of the company, *i* has the the power and the authority to command *j* to participate to the meeting. This is the reason why, after *i*'s act of delegation, *j* will be obliged to participate to the meeting.

Note that in the previous scenario, the Conditions D and E allow to distinguish order from a simple request. Moreover, F is the condition that guarantees that the speaker's act will produce normative effects on the hearer. For instance, suppose that agent *z* is another employee of the company. Again, *j* has been instructed by the president of the company to participate to the journalist meeting. But *j* has no wish to do this and decides to ask his colleague *z* a favour. Agent *j* does not have the power and authority to command *z* to participate to the meeting. Thus, agent *j* cannot order *z* to participate to the meeting and cannot delegate to *z* this task. At the most *j* can make a simple request to *z* to participate to the meeting instead of him.

Note also that the previous scenario mentions the concept of *representation*, whereby an agent *j* (the representative) can act in the name of another agent *i* (the principal), that is, *j* has the power to represent *i* at the journalist meeting (see, e.g., [11] for an account of the concept of representation in legal theory). In the scenario, representation is obtained by means of the delegation act by *i*. Moreover, note that the act of delegation can be conceived as a *mandate* whose effect is the creation of the obligation for *j* of exercising the power of representation. We postpone to future work an analysis of the relations between our notions of delegation, and the notions of representation and mandate.

Before concluding, let us observe that in certain situations one would like to call delegation an action of agent *i* directed to agent *j* even though *i* (the speaker) did not tell explicitly to *j* (the hearer) that if *j* does not perform action α , *j* will violate a norm and will be liable to punishment. We call *implicit delegation* this weaker form of delegation. In implicit delegation, *i* wants that *j* does action α , and has the power and the authority to command *j* to do α (and believes this). Moreover, *i* asks *j* to do α (by telling to *j* that *i* wants that *j* does α) but *i* does not explicitly reveal to *j* that *i*'s request will have certain normative effects on *j* (i.e. will oblige *j* to do action α). In other words, in implicit delegation, the speaker performs a simple request to someone knowing that by making the request the other will be obliged to perform the required action. For example, the president of a company may delegate a certain task to his secretary in a very polite way ('Could you please fix the appointment for me?'), with-

out explicitly ordering her to accomplish the task, but knowing that after the request the secretary will be obliged to accomplish the task (given that the president has the power and the authority to oblige the secretary to accomplish the task by asking her to do so). In the rest of analysis we will leave aside this concept of implicit delegation, and we will consider the concept of delegation in a full sense.

The rest of the paper is organized as follows. In Section 3 we introduce a modal logic of mental states, action and communication. The interesting aspect of this logic is that it allows to describe what an agent expresses in the context of conversation and to characterize the social dimension of communication (not simply its mental dimension). It has long been argued that multi-agent systems should have a social semantics (see e.g. [23, 25]), that is a semantics about externally observable objects (such as commitment, permissions, obligations, etc.) rather than about agents' internal mental states (beliefs, intentions, goals, etc.). Our point of view here is that observable states of affairs involved in communication are nothing else than the open and public expressions of internal mental states of the speaker. The logic presented in the next section will enable capture this aspect of communication in a proper way. In Section 4 the modal logic of Section 3 will be first used in order to define the concept of obligation. Then, it will be used to define request and order and, finally, the concept of delegation. In the logical definitions of request and order we will just consider the external (overt) conditions of speech acts (Conditions C and E above). Indeed, these are the most interesting conditions when looking at the social level of communication (e.g. how a speech act affects the social world by helping to achieve co-ordination, by creating obligations, and so on). Finally, we will shed light on the logical relationship between delegation and obligation, that is, we will explain why a delegation act of *i* directed to *j* can create an obligation for *j* to perform the action delegated by *i*.

3 A modal logic of mental states, action and communication

We present in this section a modal logic of agents' mental attitudes including beliefs and goals, extended by communication modalities which enable to describe what a given agent tells to another agent in front of a group of agents (an audience). In order to be able to reason about actions of agents and about speech acts of directive type like requests and orders, we include in our logic special atomic propositions denoting action occurrences.

3.1 Syntax

Assume a finite set $AGT = \{1, \dots, n\}$ of labels denoting agents, a countable set $FACT = \{p, q, \dots\}$ of atomic formulae denoting facts, and a countable set of atomic formulae denoting action occurrences $ACT = \{\alpha_1, \beta_1, \dots, \alpha_2, \beta_2, \dots, \alpha_n, \beta_n\}$. For example, if α_1 is true then it means that agent 1 performs action α , if β_2 is true then it means that agent 2 performs action β , and so on. We define $ATM = FACT \cup ACT$. Moreover, we define $2^{AGT*} = 2^{AGT} \setminus \emptyset$ the set of all non-empty sets of agents.

The language $\mathcal{L}$ of the logic **L** is the set of formulas defined by the following BNF:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathbf{Bel}_i\varphi \mid \mathbf{Goal}_i\varphi \mid \mathbf{Tell}_{i,j,C}\varphi \mid \mathbf{X}\varphi$$

where p ranges over ATM , i ranges over AGT and C ranges over 2^{AGT*} . The other Boolean constructions $\top$, $\perp$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from $\neg$ and $\wedge$ in the standard way.

Modalities $\mathbf{Bel}_i$ and $\mathbf{Goal}_i$ are standard doxastic modalities and standard modalities for goals [5]. The formula $\mathbf{Bel}_i\varphi$ has the following reading ‘agent i believes that φ ’, whereas $\mathbf{Goal}_i\varphi$ has to be read ‘agent i has the goal that φ ’.

Modalities $\mathbf{Tell}_{i,j,C}$ are used to describe what an agent told to another agent in the presence of a group of agents. We read $\mathbf{Tell}_{i,j,C}\varphi$ as ‘agent i told to agent j that φ in the presence of the agents in C ’ or ‘agent i told to agent j that φ in front of group C ’. Given a formula $\mathbf{Tell}_{i,j,C}\varphi$, we call i the *speaker*, j the *addressee*, C the *audience* and φ the *content* of the speaker’s expression (utterance). For example, formula $\mathbf{Tell}_{i_1,i_2,C_1} \text{It is sunny!}$ has to be read ‘agent i_1 told to agent i_2 that it is sunny in front of group C_1 ’. When the speaker and the addressee are the same agent, the expression $\mathbf{Tell}_{i,i,C}\varphi$ refers to a solitary speech of agent i (a monologue).

Finally, the temporal operator $\mathbf{X}$ is the standard *next* operator of linear temporal logic [8]. Formula $\mathbf{X}\varphi$ means ‘ φ will be true in the next state’. The following abbreviation defines the concept of achievement goal in a way similar to [5]:

$$\mathbf{AGoal}_i\varphi \stackrel{\text{def}}{=} \mathbf{Goal}_i\mathbf{X}\varphi \wedge \mathbf{Bel}_i\neg\varphi$$

An agent i has the achievement goal that φ (noted $\mathbf{AGoal}_i\varphi$) if and only if i has the goal that φ will be true in the next state and believes that φ is currently false. Formula $\mathbf{AGoal}_i\varphi$ can also be read ‘agent i wants φ to be true’.

3.2 Semantics

We use a standard possible worlds semantics. Models of the logic $\mathbf{L}$ ($\mathbf{L}$ -models) are tuples $M = \langle W, \mathcal{B}, \mathcal{G}, \mathcal{T}, \mathcal{X}, \mathcal{V} \rangle$ where:

- W is a nonempty set of possible worlds or states;
- $\mathcal{B} : AGT \rightarrow W \times W$ maps every agent $i \in AGT$ to a serial, transitive and Euclidean relation $\mathcal{B}_i$ between worlds in W ;
- $\mathcal{G} : AGT \rightarrow W \times W$ maps every agent $i \in AGT$ to a serial relation $\mathcal{G}_i$ between worlds in W ;
- $\mathcal{T} : AGT \times AGT \times 2^{AGT^*} \rightarrow W \times W$ maps every agent $i \in AGT$ and non-empty set of agents $C \in 2^{AGT^*}$ to a transitive relation $\mathcal{T}_{i,j,C}$ between worlds in W ;
- $\mathcal{X} : W \rightarrow W$ is a total function mapping every state in W to its successor state;
- $\mathcal{V} : ATM \rightarrow 2^W$ is a valuation function.

Accessibility relations on W can be viewed as functions from W to 2^W . Therefore, we write $\mathcal{B}_i(w) = \{v | (w, v) \in \mathcal{B}_i\}$, $\mathcal{G}_i(w) = \{v | (w, v) \in \mathcal{G}_i\}$ and $\mathcal{T}_{i,j,C}(w) = \{v | (w, v) \in \mathcal{T}_{i,j,C}\}$.

The set $\mathcal{B}_i(w)$ is the *information state* of agent i at world w : the set of worlds that agent i considers possible at world w . The set $\mathcal{G}_i(w)$ is the *preference state* of agent i at world w : the set of worlds that agent i prefers at world w . The set $\mathcal{T}_{i,j,C}(w)$ is the *communication state* of agent i with respect to agent j in the presence of group C at world w : the set of worlds that are compatible with what agent i told to agent j in front of group C at world w .

$\mathbf{L}$ -models are supposed to satisfy the following additional constraints. For every world $w \in W$, for all $i, j \in AGT$, for all $C \in 2^{AGT^*}$, if $z \in C \cup \{i, j\}$ we have:

- S1 if $v \in \mathcal{B}_i(w)$ then $\mathcal{G}_i(v) \subseteq \mathcal{G}_i(w)$;
- S2 if $v \in \mathcal{B}_i(w)$ then $\mathcal{G}_i(w) \subseteq \mathcal{G}_i(v)$;
- S3 if $v \in \mathcal{B}_z(w)$ then $\mathcal{T}_{i,j,C}(v) \subseteq \mathcal{T}_{i,j,C}(w)$;
- S4 if $v \in \mathcal{B}_z(w)$ then $\mathcal{T}_{i,j,C}(w) \subseteq \mathcal{T}_{i,j,C}(v)$.

Constraints S1 and S2 are about positive and negative introspection for goals. They are together equivalent to: if $v \in \mathcal{B}_i(w)$ then $\mathcal{G}_i(w) = \mathcal{G}_i(v)$. This means that worlds that are preferred by agent i are also preferred by agent i from those worlds that he considers possible.

Constraints S3 and S4 are about positive and negative introspection for communication, and are together equivalent to: if $v \in \mathcal{B}_z(w)$ then $\mathcal{T}_{i,j,C}(v) = \mathcal{T}_{i,j,C}(w)$, for every $z \in C \cup \{i, j\}$. Suppose that $z \in C \cup \{i, j\}$. Then, S3 and S4 together means that: worlds that are compatible with what agent i told to agent j in front of group C , are also compatible with what agent i told to agent j in front of group C from those worlds that agent z considers possible.

Given a model M , a world w and a formula φ , we write $M, w \models \varphi$ to mean that φ is true at world w in M . The rules defining the truth conditions of Boolean constructions are entirely standard. The truth conditions of modalities $\mathbf{Bel}_i$, $\mathbf{Goal}_i$, $\mathbf{Tell}_{i,j,C}$ and $\mathbf{X}$ are defined as follows:

- $M, w \models \mathbf{Bel}_i\varphi$ iff $M, v \models \varphi$ for all v such that $v \in \mathcal{B}_i(w)$;
- $M, w \models \mathbf{Goal}_i\varphi$ iff $M, v \models \varphi$ for all v such that $v \in \mathcal{G}_i(w)$.
- $M, w \models \mathbf{Tell}_{i,j,C}\varphi$ iff $M, v \models \varphi$ for all v such that $v \in \mathcal{T}_{i,j,C}(w)$;
- $M, w \models \mathbf{X}\varphi$ iff $M, v \models \varphi$ for $v = \mathcal{X}(w)$.

We write $\models_{\mathbf{L}} \varphi$ if φ is *valid* in $\mathbf{L}$ (φ is true in all $\mathbf{L}$ -models).

3.3 Axiomatization

Fig. 2 contains the axiomatization of the logic $\mathbf{L}$.

(KD45 _{Bel})	All KD45-principles for the operators $\mathbf{Bel}_i$
(KD _{Goal})	All KD-principles for the operators $\mathbf{Goal}_i$
(K4 _{Tell})	All K4-principles for the operators $\mathbf{Tell}_{i,j,C}$
(K _X)	All K-principles for the operators $\mathbf{X}$
(PI _{Goal})	$\mathbf{Goal}_i\varphi \rightarrow \mathbf{Bel}_i\mathbf{Goal}_i\varphi$
(NI _{Goal})	$\neg\mathbf{Goal}_i\varphi \rightarrow \mathbf{Bel}_i\neg\mathbf{Goal}_i\varphi$
(PI _{Tell})	$\mathbf{Tell}_{i,j,C}\varphi \rightarrow \mathbf{Bel}_z\mathbf{Tell}_{i,j,C}\varphi$ (if $z \in C \cup \{i, j\}$)
(NI _{Tell})	$\neg\mathbf{Tell}_{i,j,C}\varphi \rightarrow \mathbf{Bel}_z\neg\mathbf{Tell}_{i,j,C}\varphi$ (if $z \in C \cup \{i, j\}$)
(Funct _X)	$\mathbf{X}\varphi \leftrightarrow \neg\mathbf{X}\neg\varphi$

Figure 2. Axiomatization of $\mathbf{L}$

We have all principles of the normal modal logic KD45 for every belief operator $\mathbf{Bel}_i$. Thus, an agent cannot have inconsistent beliefs (i.e. $\neg(\mathbf{Bel}_i\varphi \wedge \mathbf{Bel}_i\neg\varphi)$), and he has positive and negative introspection over his beliefs (i.e. $\mathbf{Bel}_i\varphi \rightarrow \mathbf{Bel}_i\mathbf{Bel}_i\varphi$ and $\neg\mathbf{Bel}_i\varphi \rightarrow \mathbf{Bel}_i\neg\mathbf{Bel}_i\varphi$). We have all principles of the normal modal logic KD for every goal operator $\mathbf{Goal}_i$. Thus, an agent cannot have inconsistent goals (i.e. $\neg(\mathbf{Goal}_i\varphi \wedge \mathbf{Goal}_i\neg\varphi)$). Moreover, we have all principles of the normal modal logic K4 for every communication operator $\mathbf{Tell}_{i,j,C}$. Thus, i ’s action of telling to j that φ entails i ’s action of telling to j that i tells to j that φ (i.e. $\mathbf{Tell}_{i,j,C}\varphi \rightarrow \mathbf{Tell}_{i,j,C}\mathbf{Tell}_{i,j,C}\varphi$). In other words, the action of telling something to someone has a self-referential nature.

Axioms PI_{Goal} and NI_{Goal} are standard axioms of positive and negative introspection for goals: if an agent has the goal (resp. does not have the goal) that φ , he believes this.

Finally, $\mathbf{PI}_{\text{Tell}}$ and $\mathbf{NI}_{\text{Tell}}$ are corresponding principles of positive and negative introspection for communication. According to Axiom $\mathbf{PI}_{\text{Tell}}$, if agent i told to agent j that φ in front of an audience C then agent i , agent j and every agent in the audience C believes this. According to Axiom $\mathbf{NI}_{\text{Tell}}$, if agent i did not tell to agent j that φ in front of the agents in C then agent i , agent j and every agent in C believes this. This means that if an agent i tells (resp. does not tell) something to someone in front of an audience C then this is public in the group $C \cup \{i, j\}$ including the speaker, the addressee, and all agents in the audience.

REMARK. Note that we did not include a general inclusion principle of the form $\mathbf{Tell}_{i,j,C}\varphi \rightarrow \mathbf{Tell}_{i,j,B}\varphi$ for $B \subseteq C$ stating that if B is a subgroup of C and i told to j that p in front of C then i also told to j that p in front of B . In fact, we want to be able to model situations in which an agent i told something in secret to another agent j (while all other agents were not hearing), and he told the contrary to j in front of a large group including j . For example, Bill might tell in secret to Mary that he loves Ann, i.e. $\mathbf{Tell}_{\text{Bill},\text{Mary},\{\text{Mary}\}}\text{BillLovesAnn}$, while Bill tells to Mary that he does not love Ann when he is in front of Mary and Bob (because Bill wants that Bob does not know that he loves Ann), i.e. $\mathbf{Tell}_{\text{Bill},\text{Mary},\{\text{Mary},\text{Bob}\}}\neg\text{BillLovesAnn}$.

We call $\mathbf{L}$ the logic axiomatized by the principles given in Fig. 2. We write $\vdash_{\mathbf{L}} \varphi$ if φ is a $\mathbf{L}$ -theorem.

Theorem 1 *The logic $\mathbf{L}$ is completely axiomatized by the principles in Fig. 2.*

PROOF. It is a routine task to check that the axioms of the logic $\mathbf{L}$ correspond one-to-one to their semantic counterparts on the models. In particular, $\mathbf{KD45}_{\text{Bel}}$ corresponds to the fact that every $\mathcal{B}_i$ is a serial, transitive and Euclidean relation; $\mathbf{KD}_{\text{Goal}}$ corresponds to the fact that every $\mathcal{G}_i$ is a serial relation; $\mathbf{K4}_{\text{Tell}}$ corresponds to the fact that every $\mathcal{T}_{i,j,C}$ is a transitive relation. Moreover, $\mathbf{PI}_{\text{Goal}}$ and $\mathbf{NI}_{\text{Goal}}$ correspond respectively to the constraint S1 and to the constraint S2 on $\mathbf{L}$ -models, $\mathbf{PI}_{\text{Tell}}$ and $\mathbf{NI}_{\text{Tell}}$ correspond respectively to the constraint S3 and to the constraint S4 on $\mathbf{L}$ -models, and $\mathbf{Funct}_X$ correctly axiomatizes the successor function $\mathcal{X}$.

It is routine, too, to check that all of our axioms are in the Sahlqvist class. This means that the axioms are all expressible as first-order conditions on models and that they are complete with respect to the defined model classes, cf. [3, Th. 2.42]. ■

The following are examples of $\mathbf{L}$ theorems.

Proposition 1 *For every $i, j \in \text{AGT}$ and for every $C \in 2^{\text{AGT}^*}$ we have:*

- $$(1a) \quad \vdash_{\mathbf{L}} \mathbf{Tell}_{i,j,C}\varphi \leftrightarrow \bigwedge_{z \in C \cup \{i,j\}} \mathbf{Bel}_z \mathbf{Tell}_{i,j,C}\varphi$$
- $$(1b) \quad \vdash_{\mathbf{L}} \neg \mathbf{Tell}_{i,j,C}\varphi \leftrightarrow \bigwedge_{z \in C \cup \{i,j\}} \mathbf{Bel}_z \neg \mathbf{Tell}_{i,j,C}\varphi$$

According to Theorem 1a, agent i told to agent j that φ in front of the audience C if and only if i, j and every agent in the audience believes this. According to Theorem 1b, agent i did not tell to agent j that φ in front of the audience C if and only if i, j and every agent in the audience believes this.

4 A logical analysis of delegation

In the rest of the paper we will use the modal logic $\mathbf{L}$ in order to provide a logical analysis of delegation. We will first use this logic in or-

der to define some normative concepts like the concept of obligation-to-be and the concept of obligation-to-do (Section 4.1). Then, we will use it in order to define request and order (Section 4.2). Finally (in Section 4.3), we will first define the concept of power to command and the concept of delegation. We will conclude with a clarification of the logical relationships between delegation and obligation, that is, we will explain why a delegation act by i towards j will create an obligation for j to perform the action delegated by i .

4.1 Some normative concepts

Let us introduce a nonempty finite set of symbols denoting social roles $\text{ROLE} = \{x, y, \dots\}$. The function

$$\mathcal{F}_{\text{play}} : \text{ROLE} \longrightarrow 2^{\text{AGT}^*}$$

maps every social role to a nonempty set of agents, and $\mathcal{F}_{\text{play}}(x) = C$ just means that C is the set of agents in the society playing role x . In the set ROLE we distinguish a specific role called *Auth*. If $i \in \mathcal{F}_{\text{play}}(\text{Auth})$, it means that agent i plays the role of legal authority in the society. Legal authorities are those agents that provide the criteria for the existence and the validity of a given norm in a society. In particular, a given norm is valid in the society if and only if the legal authorities recognize it as a norm. The term legal authority is for us a very general term that in certain situations is synonymous of the terms leader or legislator. For example, a contract stipulated by two agents is valid (and consequently the obligations stemming from it are valid) if and only if the legal authority (the notary) recognizes it as a valid contract.

We are interested here in defining a notion of norm based on its recognition by legal authorities. To this aim, we first define the following concept of official belief:

$$\mathbf{OffBel} \varphi \stackrel{\text{def}}{=} \bigwedge_{i \in \mathcal{F}_{\text{play}}(\text{Auth})} \mathbf{Bel}_i \varphi$$

We say that φ is officially believed or officially recognized to be true (noted $\mathbf{OffBel} \varphi$) if and only if, all legal authorities believe that φ is true. From the previous concept of official belief, we can define the concept of obligation based on the recognition by legal authorities.

$$\mathbf{Oblig} \varphi \stackrel{\text{def}}{=} \mathbf{OffBel} (\neg \varphi \rightarrow V)$$

We say that φ is obligatory (noted $\mathbf{Oblig} \varphi$) if and only if it is officially believed that the states in which φ is false are *bad* states, where V is a special propositional atom used for labeling sub-ideal states (also called violation states) as in Anderson's reduction of deontic logic to alethic logic [1]. The concept of obligation-to-do is then defined in the standard way as follows:

$$\mathbf{Oblig}_i \alpha \stackrel{\text{def}}{=} \mathbf{Oblig} \alpha_i$$

We say that agent i has the obligation to do action α (noted $\mathbf{Oblig}_i \alpha$) if and only if it is obligatory that agent i does action α .

4.2 Definitions of request, order and delegation

As emphasized in Section 2, the concept of delegation is tightly connected with the concepts of request and order. So, an analysis of delegation calls for a clarification of the concepts of request and order.

As anticipated above, we just consider here the external (overt) conditions of speech acts, that is, the conditions referring to what

the speaker overtly expresses while performing the speech act. In particular, we define request and order on the basis of Condition C and Condition E given in Section 2.

So, we say that agent i made a request to agent j in front of an audience C to do action α (noted **Request**(i, j, C, α)) if and only if, i told to j in front of the audience C that i wants that j does action α :

$$\text{Request}(i, j, C, \alpha) \stackrel{\text{def}}{=} \text{Tell}_{i, j, C} \text{AGoal}_i \alpha_j$$

Moreover, we say that agent i ordered agent j in front of an audience C to do action α (noted **Order**(i, j, C, α)) if and only if, i told to j in front of the audience C that i wants that j does action α and that if in the next state j does not perform this action then this counts as a violation of a norm by j :

$$\text{Order}(i, j, C, \varphi) \stackrel{\text{def}}{=} \text{Tell}_{i, j, C} (\text{AGoal}_i \alpha_j \wedge (\mathbf{X} \neg \alpha_j \rightarrow \mathbf{V}))$$

To simplify our exposition, in the rest of the paper we will omit the third argument concerning the audience in the two predicates for request and order when $C = \mathcal{F}_{\text{play}}(\text{Auth})$. That is, we will simply write **Request**(i, j, φ) instead of **Request**($i, j, \mathcal{F}_{\text{play}}(\text{Auth}), \varphi$), and **Order**(i, j, φ) instead of **Order**($i, j, \mathcal{F}_{\text{play}}(\text{Auth}), \varphi$). Therefore, **Request**(i, j, φ) and **Order**(i, j, φ) are simply read “agent i made a request to agent j to ensure φ ” and “agent i ordered agent j to ensure φ ”.

4.3 Creation of obligations through delegation

We conclude our formal analysis by clarifying the relationship between delegation and obligation, that is, by explaining how a delegation act by i directed towards j will create an obligation for j to perform the action delegated by i .

The following notion of power to command is a fundamental building block of our definition of delegation:

$$\text{CPower}(i, j, \alpha) \stackrel{\text{def}}{=} \text{OffBel}((\text{Order}(i, j, \alpha) \rightarrow (\mathbf{X} \neg \alpha_j \rightarrow \mathbf{V}))$$

We say that agent i has the power to command agent j to do α (noted **CPower**(i, j, α)) if and only if the legal authorities recognize that if agent i ordered agent j to do action α and in the next state j does not perform this action then, this counts as a violation of a norm by j .

Delegation is defined by means of these concepts of order and of power to command. In particular, we say that agent i delegated action α to agent j (noted **Deleg**(i, j, α)) if and only if i ordered j to do action α and i has the power to command j to do action α :

$$\text{Deleg}(i, j, \varphi) \stackrel{\text{def}}{=} \text{Order}(i, j, \varphi) \wedge \text{CPower}(i, j, \alpha)$$

The following theorem is a fundamental contribution of the present analysis since it sheds light on the hidden path from delegation to obligations.

Proposition 2 For every $i, j \in \text{AGT}$ we have:

$$\vdash_L \text{Deleg}(i, j, \alpha) \rightarrow \text{Oblig}_j \alpha$$

⁵ Note that i 's power to command agent j to do action α should be conceived as a specific form of institutional power based on so-called constitutive rules [21] recognized to be valid by the legal authorities of the form ‘ i 's act of commanding agent j to do action α counts as i 's act of ensuring that j has the obligation to accomplish that action’ (see, e.g., [17, 15, 12] for some logical accounts of *counts as*). The relationship between constitutive rules and delegation power will not be analyzed here.

The previous L-theorem just says that if i delegated action α to j (on the basis of the recognition by the legal authorities of i 's power to command j to do action α) then, as a consequence, j has the obligation to do action α . Let us go back to the example given in Section 2. If agent i_1 , the president of a company, delegated to agent i_2 , an employee of the company, the task of representing the company at a journalist meeting (i.e. **Deleg**($i_1, i_2, \text{reprCompany}$)) then, necessarily, i_2 has the obligation to represent the company at the journalist meeting (i.e. **Oblig** _{i_2} *reprCompany*), since the legal authorities recognize i_1 's power to command i_2 to represent the company at the journalist meeting.

Before concluding, let us add a short remark. We are aware that there are cases of delegation in which the delegant's power to command is a form of absolute power which does not need to be legitimated by external legal authorities. This is for instance the case of the absolute power of a sovereign or of a general of an army. One might say that in this case, the legal authority and the agent to whom the power to command is conferred coincide, that is, we have **CPower**(i, j, α) and $\mathcal{F}_{\text{play}}(\text{Auth}) = \{i\}$. For example, it is the general himself who legitimates his power to command his soldier to do a certain action α , i.e. **CPower**(*general*, *soldier*, α) and $\mathcal{F}_{\text{play}}(\text{Auth}) = \{\text{general}\}$. Note that, following the Hobbesian tradition, this form of absolute power is in turn founded upon an act of delegation by the members of the society (also called delegation of power). We postpone to future work an analysis of this form of delegation of power to a supreme authority.

5 Related works

Several authors working in the area of intersection between social sciences and computer science have been interested in the concept of delegation and have tried to clarify the role of delegation in organized social action. We here discuss some of the most important approaches to delegation.

Falcone & Castelfranchi's model of delegation. Falcone & Castelfranchi [9, 10] have proposed a model of delegation that has some similarities with the approach proposed in this work. They are interested in a specification of the cognitive constituents of delegation as a particular cognitive state (defined in terms of beliefs, goals, intentions) of the delegating agent. Castelfranchi & Falcone individuate several kinds of delegation. They introduce a first type of delegation called *weak delegation*, that is, the delegation based on exploitation, on the passive achievement by an agent i of a certain task. In weak delegation, an agent i just exploits in his plan a fully autonomous action of another agent j . In fact, i has only to recognize the possibility that j will realize a certain action and that this action ensures that the goal of i will be satisfied. In this case agent i ‘passively’ awaits the satisfaction of his goals given his expectation that j will ensure it. Then, they have a second type of delegation called *mild delegation*, the delegation based on the active indirect achievement by i of the task. Agent i not only exploits agent j 's action for the achievement of his goals since he sees agent j 's intervention as fundamental for this, but also acts in order to induce j to perform the right course of action. Finally, they have a third type of delegation called *strong delegation* based on goal adoption. In strong delegation, agent i has a certain goal φ and decides to act in order to induce j to adopt his goal φ . Although Castelfranchi & Falcone recognize that strong delegation might be based on a request or order of the delegant to the delegate, they leave implicit the communicative aspect of delegation and do not consider in detail the relationships between delegation and the speech acts request and order. Finally,

Castelfranchi & Falcone's analysis does not consider the relationships between delegation, power to command and obligations which is a fundamental point in our analysis. That is, Castelfranchi & Falcone do not consider how i 's act of delegating to j an action α will create an obligation for j to perform the action, given the recognition by the legal authorities of i 's power to command j to do action α .

'Non-mentalistic' approaches to delegation. Recently other researchers have developed formal systems for reasoning about delegation. Many of these approaches try to express the concept of delegation without referring to mental states. Norman & Reed [18] analyze those forms of delegation where by issuing an imperative an agent i creates an obligation for another agent j to accomplish a given task. They exploit a logic of agentive action by using a *bringing it about* modal operator E_i similar to [15] and a deontic modality O . According to them a formula such as $E_i O E_j \varphi$ reading "agent i brings it about that the state of affairs holds in which it is obligatory for j to see to it that φ " corresponds to i 's action of issuing an imperative to j and expresses delegation. Santos et al. [19] proposes a similar analysis of delegation (or attribution) of responsibility. A similar approach to delegation is also proposed by Grossi et al. [13] in which delegation is taken as primitive action without being analyzed in terms of the cognitive states of the delegant, and what delegation does is just creating directed obligations, given that certain necessary preconditions for the creation of the obligation hold.

Although we agree with Norman & Reed [18], Santos et al. [19] and Grossi et al. [13] that delegation is responsible for the creation of obligations (see Section 4.3), we think that it is too simplistic to reduce delegation to something like "agent i brings it about that the state of affairs holds in which it is obligatory for j to see to it that φ " as Norman & Reed and Santos et al. do. The main problem with this approach is that it is not able to exclude from the definition of delegation all those situations where by performing a certain action, an agent i *accidentally* (*non intentionally*) brings it about that agent j is subject to a certain obligation. For example, by passing through the security gate at the airport with a suspicious hand luggage, a terrorist named Oscar it about that the state of affairs holds in which it is obligatory for the security service to search Oscar's hand luggage (suppose that there is a general rule saying that "security service must search all suspicious hand luggage"): $E_{Oscar} O E_{Security} Search Oscar Luggage$. It seems quite odd to say that Oscar delegates the security service to search his hand luggage. Indeed, Oscar neither wants nor intends that the security service will search his hand luggage. It is rather the authority of the airport that having the intention to prevent terrorist attacks delegates to the security service the general task of searching all suspicious hand luggage and the more specific task of searching Oscar's luggage.

In our view in order to provide an adequate analysis of the concept of delegation, we must seriously consider both its *intentional dimension* and its *communicative dimension*. As emphasized in the present analysis, we can say that an agent i delegates a task α to another agent j only if i wants that j will accomplish this task, i communicates to j his intention, and i communicates to j that if he does accomplish the task he will incur the violation of a norm. The previous 'non-mentalistic' approaches completely ignore these two crucial dimensions of the concept of delegation.

6 Conclusion

We have presented in this work a conceptual analysis of delegation. We have conceived delegation as a social action involving two agents called the delegant and the delegate that the delegant performs by giv-

ing an order to the delegate, and that produces certain normative effects on the delegate and, in particular, an obligation for the delegate to perform the delegated task. An aspect that we intend to investigate in the future is the relationship between trust and delegation. Indeed, an agent i 's decision to delegate some task to another agent j is often based on i 's trust in j , that is, in many situations trust is a necessary precondition for delegation. To this aim, we intend to integrate the analysis presented here with our previous works on trust [16].

REFERENCES

- [1] A. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 22:100–103, 1958.
- [2] K. Arrow. *The Limits of Organization*. Norton, New York, 1974.
- [3] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge Univ. Press, 2001.
- [4] M. Bratman. *Intentions, plans, and practical reason*. Harvard Univ. Press, Cambridge, 1987.
- [5] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.
- [6] J. Coleman. *Foundations of Social Theory*. Harvard Univ. Press, Cambridge, 1990.
- [7] F. Dignum. Autonomous agents with norms. *Artificial Intelligence and Law*, 7:69–79, 1999.
- [8] E. A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*. MIT Press, 1990.
- [9] R. Falcone and C. Castelfranchi. Towards a theory of delegation for agent-based systems. *Robotics and Autonomous Systems*, 24:141–157, 1998.
- [10] R. Falcone and C. Castelfranchi. The human in the loop of a delegated agent: the theory of adjustable social autonomy. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Sysems and Humans*, 31(5):406–418, 2001.
- [11] J. Gelati, A. Rotolo, G. Sartor, and G. Governatori. Normative autonomy and normative co-ordination: Declarative power, representation, and mandate. *Artificial Intelligence and Law*, 12((1-2)):53–81, 2004.
- [12] D. Grossi, J.-J. Ch. Meyer, and F. Dignum. Classificatory aspects of counts-as: An analysis in modal logic. *Journal of Logic and Computation*, 16(5):613–643, 2006.
- [13] D. Grossi, L. Royakkers, and F. Dignum. Organizational structure and responsibility: an analysis in a dynamic logic of organized collective agency. *Artificial Intelligence and Law*, 15:223–249, 2007.
- [14] H. Herrestad and C. Krogh. Obligations directed from bearers to counterparties. In *Proc. of the Fifth International Conference on Artificial intelligence and law*, pages 210–218. ACM Press, 1995.
- [15] A. Jones and M. J. Sergot. A formal characterization institutionalised power. *Logic Journal of the IGPL*, 4:429–445, 1996.
- [16] E. Lorini and R. Demolombe. Trust and norms in the context of computer security. In *Proc. of DEON'08*, pages 50–64. Springer-Verlag, 2008.
- [17] E. Lorini and D. Longin. A logical account of institutions: from acceptances to norms via legislators. In *Proc. of KR 2008*, pages 38–48. AAAI Press, 2008.
- [18] T. J. Norman and C. A. Reed. A model of delegation for multi-agent systems. In *Foundations and Applications of Multi-Agent Systems*. Springer Verlag, 2002.
- [19] F. Santos, J. Carmo, and A. Jones. Action concepts for describing organised interaction. In *Thirtieth Annual Hawaii International Conference on System Sciences*, pages 373–382. IEEE Press, 1997.
- [20] J. Searle. *Speech acts: An essay in the philosophy of language*. Cambridge Univ. Press, Cambridge, 1969.
- [21] J. Searle. *The Construction of Social Reality*. The Free Press, New York, 1995.
- [22] P. Selznick. Foundations of the theory of organization. *American Sociological Review*, 13:25–35, 1948.
- [23] M. Singh. A social semantics for agent communication languages. In *Issues in Agent Communication*, pages 31–45. Springer-Verlag, 2000.
- [24] D. Vanderveken. *Principles of language use*, volume 1 of *Meaning and Speech Acts*. Cambridge Univ. Press, 1990.
- [25] M. Verdicchio and M. Colombetti. A logical model of social commitment for agent communication. In *Proc. of AAMAS 2003*, pages 528–535. ACM Press, 2003.

Power Laws and Levels of Abstraction in Understanding Legal Systems

Ugo Pagallo^{1,2}

Abstract. This paper focuses on some misunderstandings of today's debate in network analysis and the law, resulting from a lack of clarity about the 'levels of abstraction' we are dealing with. By distinguishing three levels, *i.e.*, information *as* reality, *for* reality, and *on* reality, the paper aims to prevent some of these misunderstandings: On one hand, there is a remarkable shift in current applications of social networks to legal system-analysis, namely, from traditional researching information *as* reality to current researching information *on* reality. On the other hand, we have to pay attention to the different ways in which scholars approach 'reality' – in, *e.g.*, the realm of law – because every theory commits itself ontologically by opting for a specific level of abstraction.

1 INTRODUCTION

Last year, at the Heriot-Watt University in Edinburgh, I presented a paper on the "small world"-paradigm and the topology of complex legal systems, following up on the proposal of the AISB'09 Convention, *i.e.*, to promote the interchange of knowledge and methodologies among some of the most interesting and promising fields of current scientific research, such as social network-perspectives and multi-agent systems-analyses. Along with the paper of Jeremy Pitt, colleagues and organizers honoured us awarding our work as the best paper. Therefore, I would like to show that I took stock of what I learnt by deepening the topic in a different light: I suggest to examine today's debate on the application of social networks to legal system-analysis paying attention to the methodology of the levels of abstraction (LoA) put forward by Luciano Floridi [1].

In a nutshell, a LoA can be grasped as an 'interface' that is made of a set of features, namely, the 'observables.' Given some values, an observable of a type X is a variable whose values range over that very type. Hence, a LoA consists of a collection of observables of type X, that is defined by how we choose to describe, examine, and argue about a system and its context. For example, in the realm of 'ontologies,' Guarino and Giarretta have pointed out different kinds of analysis and, therefore, the multiple meanings related to this research [2, 3]:

- i) A philosophical discipline which is concerned with being *qua* being;
- ii) An informal conceptual system;
- iii) A formal semantic account of that system;
- iv) A specification of its possible conceptualizations;
- v) The representation of a conceptual system through logical theories;
- vi) The vocabulary employed by that logical theory;
- vii) A meta-level specification of the logical theory.

As you can see, each LoA enables an analysis of the system, the result of which is a model of it: While a range of LoAs can describe a given system, that system can have several models as well. Going back to the application of social networks to legal system-analysis, it is therefore important to determine the "ontological commitment" of every approach in opting for a specific LoA [1]. There is, indeed, network analysis on:

- i) How information is produced and distributed in legal systems;
- ii) How it is feasible to harness these properties so as to 41

obtain useful applications in the legal domain;

- iii) How knowledge and concepts frame the representation and function of shared legal terminology.

This differentiation is particularly relevant because many misunderstandings of the current debate depend on the lack of clarity about the LoAs we are dealing with. By distinguishing three levels, namely, information *as* reality, *for* reality, and *on* reality, the aim of this paper is thus to prevent some of those misunderstandings via an overall picture of today's work in network thinking and its diverse "ontological commitments." More specifically, the paper is presented in five sections.

In section 2, I take into account Newman's *Power laws, Pareto distributions and Zipf's law* [4]: This article frames the theoretical premises of my work so as to understand what is going on in the field of network thinking.

In section 3, I examine how analyses on power laws-distribution of information have extended to a number of different fields like, say, the realm of music [5, 6].

In section 4, I focus on some recent work on legal networks that shed further light on the small world-paradigm [7, 8].

In section 5, I consider "problems and perspectives for knowledge discovery in the legal domain," so that "with a specific regard to legal ontologies in multicultural contexts, new advances in terms of social network analysis could be fruitful" [9].

Finally, in section 6, I stress a shift in current applications of social networks to legal system-analysis, that is, from traditional researching information *as* reality to current researching information *on* reality [10].

The conclusion is that we should pay attention to the different ways in which scholars approach 'reality' (here, the realm of law) because every theory commits itself by opting for a specific level of abstraction.

2 POWER LAWS

In his remarkable work published on *Contemporary Physics* [4], M. E. J. Newman summed up current research on the statistical distribution of quantities and, more specifically, in the so-called power-law distributions. By illustrating the mathematical foundations of the research as well as its theoretical constraints, Newman highlighted an extraordinary set of phenomena that obeys this kind of law.

All in all, the statistical measurement of quantities follows a power-law when the "vast majority of the measurements assume relatively small values," and "there is a part of the population that assumes significantly larger values than the usual measurements" [5]. In other words, power-law distribution is equivalent to both Zipf's law and Pareto distribution, the only difference consisting in the way the cumulative distribution of information is plotted. This means that, notwithstanding the model, "the data depicted in the plots are identical" [4].

Newman examined twelve examples of power-law distributions: Word frequency, citations of scientific papers, web hits, copies of books sold, telephone calls, magnitude of earthquakes, diameter of moon craters, intensity of solar flares and of wars, wealth of the richest people, and so forth. Although "few real-world distributions follow a power law over their entire range, and in particular not for smaller values of the variable being measured," what is striking is that "one can, without stretching the interpretation of the data unreasonably, claim that power-law distributions have been observed in

¹ Law School, Univ. of Turin, Italy. Email: ugo.pagallo@unito.it

² CTLS Faculty, London.

language, demography, commerce, information and computer sciences, geology, physics and astronomy, and this on its own is an extraordinary statement” [4].

Of course, there are distributions that do not follow a power law. This is the case of the abundance of North American bird species, the number of entries in people’s email address books, or the distribution of the sizes of forest fires. Furthermore, it is quite problematic to determine the mechanisms according to which this form of distribution arises both in natural and social systems, that is, via combinations of exponentials, inverses of quantities, random walks, the “Yule process,” phase transitions and critical phenomena, self-organized criticality, etc.

Yet, it is significant that a number of extremely diverse range of phenomena can be explained in terms of power-laws of distribution. Whereas this kind of distribution means that only few nodes of the network have very high values of connectivity while most nodes present low degree, such a law has become particularly popular as the ‘long tail’ [11], or summed up as the ‘rich gets richer’-effect [12]. Because such success entails its own risks, let me further illustrate what is at stake with the peculiar case of music.

3 A CONTEXT OF MUSIC

Research on power-laws distribution in musical data is not new at all. In 1975, Voss and Clarke claimed that both pitch and loudness fluctuation in several kinds of music as classical, jazz, or rock, obeyed Zipf’s law [13]. Later, in Hsu *et al.*, it was noted that acoustic frequency in a collection of Bach changes according to a scale free distribution of musical information [14]. Following Newman’s analysis on the panoply of different phenomena characterized by power-law distributions, recent work has shed further light on these laws.

For example, by considering some quantities like the “number of musical scores that a repeating pattern is contained in” (*support*), the “number of notes contained in the repeating pattern” (*length*), and the “number of occurrences of a repeating pattern in the same musical score” (*frequency*), Rafailidis and Manolopoulos argue that music belongs to the numerous phenomena in which information is distributed according to power-laws [5]. Based on the Kolmogorov-Smirnov test, they show that this is the case of some famous compositions of Bach, Corelli, and Haydn. (The musical pieces are available at <http://kern.humdrum.org> in the kern and midi format.) This perspective would thus allow us to further understand “the differences among diverse music collections,” together with music classification and the mathematical skills for developing systems that “could compose music based on statistical properties” [5].

However, it has been stressed that music “does not possess any conventionally defined units of meaning” [6]. Thereby, in order to single out the semantic features of the subject, we should take into account the position and duration of the single notes in the light of the process of a text generation. This approach is similar to the way in which Herbert Simon updated Zipf’s law by examining the probability of using a word while writing a text [15]. The analysis of the Musical Instrument Digital Interface or MIDI files available at www.geocities.com/benedetto_marcello/midi/, has proved that the probability to choose a note is proportional to that note’s recurrence in a text. Since notes not yet appeared have non-zero probability of being added, the result of this process is compatible with the predictions of Simon’s model: “While the extension of the notion of semantic contexts from linguistics to music holds as a metaphoric allegory only, context – whose role in language is closely related to semantics – stands for a significant feature common to linguistic and musical messages. In both domains, context denotes a property emerging from the interaction of the perceptual elements that compose the

message, that makes the message intelligible as a whole. (...) [T]he combination of those elements in a hierarchically organised sequence, whose structure sustains its comprehensibility, lies at the basis of the creation of context in the two domains” [6].

For centuries, scholars abandoned the classical topic of music and the law, that is, work on their structural analogies. Without discussing the ideas of the Pythagoreans or Jean Bodin, the quantitative parallelism between musical and linguistic contexts suggests we should further analyse this connection in the light of some recent research in law, power-laws, and small world-networks. This work allows us to insist on the different levels of abstraction which are involved in the study of power-laws of distribution. As with network analysis on music, and leaving aside systems that “could compose music” by exploiting the statistical properties of the network [5], we should distinguish two different approaches of network analysis: That is, information *as* reality (e.g., support, length, and frequency in music) and information *on* reality (e.g., a musical context).

4 LAW, POWER-LAWS, AND SMALL-WORLD NETWORKS

Following my AISB’09 paper, two new results on the topology of complex legal systems have emerged in connection with the ways in which information is distributed through the nodes of a given network.

The first study focuses on the comparison of various network representations of the US Supreme Court Corpus from 1791 to 2005 [7]. As in previous research [16, 17], the body of US Supreme Court decisions, available at <http://supreme.justia.com/>, is represented as a network: Cases, or individual opinions within cases, are the nodes of the network, and citations are the links or edges of the network. The goal is to “identify the specific information captured by each representation,” to “evaluate the theoretical rationale and utility of each process,” and to give details “for parsing the plain text corpus into each network representation” [7].

Similarly to a citation network, the US Supreme Court jurisprudence can be represented as the network of connections between its cases. Yet, cases often include multiple (and dissenting) opinions, so that, in a number of relevant cases, we need a more detailed picture of the system by developing an opinion-to-opinion citation network. For example, in *Vieth v. Jubelirer* (541 U.S. 267 from 2005), we have Scalia’s decision, Kennedy’s concur opinion, and three different dissents by Stevens, Souter, and Breyer. Moreover, four of the Justices also referred to *Baker v. Carr* (269 U.S. 186 from 1962). It is clear that the five distinct units of the opinion network with its four citations would collapse into a single edge of a case-to-case network. Hence, it is necessary to increase the granularity of the analysis moving beyond citations, so as to properly distinguish the citation network from what Bommarito *et al.* call the semantic network of the US SC Corpus. The next step would be to discover how they interact: Is the web of law “seamless,” so that the categorization of legal doctrine into discrete fields like torts, contracts, etc., would not capture the nature of the law?

A fruitful way to address the issue would be to look at “the set of cases that feature both semantic similarity and a direct citation linkage. In general, these are likely communities of well-defined topical domains. Of greater interest to an empirical evaluation of the law as a seamless web, is likely the magnitude and composition of the *Citation Only* and *Semantic Only* subsets. Subject to future empirical investigation, we believe the *Citation Only* components of the graph may represent the exact type of concept exportation to and from particular semantic domains that would indeed make the law a seamless web” [7].

Awaiting the results of this new research, there is another recent study in law and network analysis that may help to

clarify the different levels of abstraction involved in this kind of work. In the network analysis of the French environmental code, Boulet *et al.* show that this code presents small-world features that clearly “contrast with the architecture found in the analysis of the reference network at the scale of codes,” that is, “a rich-club of ten codes very central to the whole French legal system, but with no small-world property.” Consequently, this level of abstraction indicates that “changing the granularity of the analysis modifies the nature of network structure and its associated properties. The density variation is not unexpected because while increasing the granularity (that is grouping vertices), the number of vertices decreases and the probability of existence of a citation between the two any bigger vertices increases” [8].

The conclusion, however, involves not only the underlying structures and properties of the law as dependent on the scale and granularity of the analysis. According to the French scholars, the small world-approach has a semantic meaning as well for it sheds light on “hidden communities” of articles in the French environmental code [8]. Whereas Boulet *et al.* propose to deepen the comparison between different environmental codes and codes related to other legal fields, their research suggests that we should to distinguish properly between two levels of abstraction, namely, between the “underlying structure” of a given legal system and the existence of “hidden communities” of articles. In the first case, we are dealing with information *as* reality; in the second case, as with Bommarito *et al.* [7]’s research on the semantic features of the US SC network, Boulet *et al.* are focusing on information *on* reality. Let us look at why this difference is so relevant.

5 KNOWLEDGE DISCOVERY

In their work on epistemological completeness of legal ontologies and knowledge discovery in the legal domain, Ajani and Rossi examined social network analysis in order to assess whether they “may contribute to integrate currently used legal ontologies” [9]. Although this approach can be useful to establish the epistemological completeness of the ontological mapping – so as to determine the effectiveness of knowledge discovery in the legal domain – a network approach would present three difficulties. We can summarize these criticisms in the following way:

First, it is argued that the topological relations examined in a small-world or power-law perspective are “made of connections among physical systems” [9]. Hence, while ontologies can be used to represent social networks, the reverse is not necessarily true. In fact, social networks would consist in representations of social interactions, where the nodes are the social actors, rather than concepts, linked by the edges of the network. From this viewpoint, the problem would be to encompass legal concepts and relations among social actors (as in the attempt of folksonomies).

Secondly, it is stressed that social network analysis for legal ontologies “cannot correspond to a representation of the whole social reality in which a top level ontology may be grounded” [9]. It follows that it is always debatable which sets of objects lie within a given network, along with network sampling and the abovementioned degree of granularity used in the analysis. “In other words, the research focus of social network analysis is too tailored on localised objectives, while conceptualisations from legal culture require a wider research focus according to the multi-layered issues of the legal domains at the international, national and regional levels” [9].

Thirdly, when small world functions are taken into account, the problem is that the focus is mainly on structural information such as the centrality, closeness, in-betweenness, etc., of the nodes that, nonetheless, would be devoid of meaning. Links or edges of the network, in fact, would only represent the relation

among nodes or subjects of the network, missing the typology of that relation. “This argument is strictly related to the rigid nature of a node-based network. Nodes are a sort of abstraction in the network and they have finite boundaries that distinguish them from their surroundings. Meaning nuances may fail to be detected when social network analysis is applied directly to semantics” [9].

At the end of the day, would not a model based on the ‘family air’ of such different fields like telephone calls and magnitude of earthquakes, word frequency and web hits, citations of scientific papers, law and music, be oversimplified?

Even though “empirical analysis by small world functions could be applied orthogonally to defined tasks, linguistic occurrences or instances with the purpose of bridging the different extensional semantics to a specific period of time and a specific social context,” the method would be “reductionist in the axiom formalisation and static in its implementation” [9].

6 LEVELS OF ABSTRACTION

When comparing multiple small-world techniques, power-law distributions, scale-free approaches, and so forth, there are limits and problems afflicting network analysis and its applications to complex social systems (such as the law). Not only power-law distributions can result from different mechanisms [18], but the very idea of ‘small’ of the small world-paradigm seems highly context-dependent: “for instance, with regard to infectious diseases, the whole world could be small – but with regard to solidarity, and mutual support, six degrees might be a whole universe apart” [19].

Nonetheless, it is important to understand what network perspectives are doing in the specific context of a given research. The aforementioned studies in the realm of law [7, 8, 16, 17], *all* deal with *how* information is distributed in complex legal networks and, yet, we need to agree on *what* ‘information’ actually means, as well as on the kind of connection between information and reality that we are examining.

First and foremost we need to go back to the difference between ‘information *as* reality’ and ‘information *on* reality’: While the former concerns the traditional field of philosophical inquiries on ontology, the latter concerns semantic information. (I do not discuss ‘information *for* reality’ here, since it exploits the topological properties of a given network so as to obtain *sets of rules or instructions* for the determination of other informational objects.)

Once we distinguish such different levels, you may notice that network analyses and the topological approaches to complex social networks mentioned in this paper mostly deal with the level of information *on* reality. Even though network analysis is traditionally focused on the statistical distribution of quantities, it is possible to grasp the topological properties of the network so as to further understand *knowledge* and *concepts* that frame the representation and function of a shared (legal) terminology. It is thus extremely important to differentiate such levels, when discussing limits and problems of the network approach to the “semantic subset” of the US SC Corpus [7], to the “hidden communities” of the French environmental code [8], and so on.

Indeed, a method does not become “reductionist” as suggested by Ajani and Rossi [9] regarding to “the axiom formalisation” of the network approach, lest information *as* reality, namely, information defining being *qua* being, is kept indistinct or confused with information *on* reality, *i.e.*, statements that involve knowledge acquired and organized by legal experts. Such a confusion takes place if legal network analysis is depicted as implying an unspoken definition of the nature of law, whereas it does not have any such claim. Rather, what is at stake in the distribution of the information represented through network analysis is the semantic

commitment to the true (or false) conformation of the real world. Contrarily to those who believe that nodes have no intrinsic meaning and links represent simple relations without defining any kind of typology, topological descriptions of complex social networks define the meaning of that very information through different scales and granularity of the analyses, multiple mechanisms for determining the laws of distribution, etc., thus enabling a comparison of their (true or false) results with the opinion of legal experts [20].

Alas, it is arguably hard to conceive a topological description of the “entire social reality,” where “meaning nuances [would] fail to be detected when social network analysis is applied directly to semantics” [9].

Let aside the description of the complete social reality, the real problem concerns the level of abstraction that is implied when comparing approaches and disciplines. Along with their multiple methodologies, there is in fact a crucial point which is common to social network-perspectives, legal ontologies, and even current research in multi-agent systems: We could imagine the preliminary assumption of every theory as a computer interface which enables both the analysis of a system and its modelling.

As I mention in the introduction, an interface is made of a collection of ‘levels of abstraction’ which includes a finite but non-empty set of observables. These ‘observables’ represent the building blocks of the theory, while defining the ‘gradient of abstractions’ or ‘scale and granularity’ of the analysis. As stressed by Floridi, an ‘observable’ should not be understood as an ‘empirically perceivable’ entity but, rather, as an ‘interpreted typed variable,’ that is, “a typed variable together with a statement of what feature of the system under consideration it stands for” [1]. While this method allows the analysis of systems by means that are necessarily developed at specific gradients of abstraction, it is the level of abstraction that enables the theory both to evaluate the system under analysis and to elaborate a model that identifies some properties of the system at the given LoA.

In the case of LoAs of network analysis, it is then important to distinguish the “ontological commitment” of this approach in opting for a specific LoA and the information captured *on* reality or processed *for* reality.

On one hand, you may indeed understand network analysis in terms of information as opposed to knowledge and concepts: *i.e.*, information *as* reality as a matter of statistical distribution of quantities in a given network.

On the other hand, information can also be exploited in order to develop some applications as set of rules or instructions to determine other informational objects: information *for* reality.

Nonetheless, there is also information *on* reality, namely, the semantic commitment of legal network analysis which is at stake in this paper with knowledge acquired and organized by scholars in both the development and evaluation-phases established by most ontology building methodologies nowadays. What about network analysis on case studies and jurisprudence [7, 16, 17], codes [8, 21], or general theory of law [22]?

All in all, integration between network theory and other areas of AI & Law does not hinge only on information *as* reality, that is, on the assumption that basic human concepts would be wired in our brains thus grounding cultural evolution. In fact, problems and limits of this approach involve the representation of a shared legal knowledge as well. If the latter may be tuned to reality in terms of evolution, it should not be confounded with the ontological quests of philosophers on being *qua* being in terms of information [23, 24].

7 CONCLUSIONS

We examined some recent research in complex social networks

and three criticisms to this approach in order to highlight a crucial point in the interchange of knowledge and methodologies among some of the most interesting and promising fields of today’s scientific research. This crucial point has been represented by the different levels of abstraction that are involved in that interchange of knowledge and methodologies proposed by the AISB meetings.

Network thinking is often represented as if it were semantically uncommitted, so that it would deal with ‘information’ as opposed to ‘knowledge’ and ‘concepts.’ Consequently, the level of abstraction at which, say, network theory and legal ontologies communicate is necessarily given by a standpoint such as Pinker’s evolutionary psychology [25].

However, it is highly debatable that network theories are invariably “made of connections between physical systems” [9], and hence, dealing with social networks like the law, it is questionable whether nodes should only represent social actors linked in the network, without any consideration of the ‘concepts’ that are involved therein.

As shown by this paper, even recent studies on power law-distributions imply multiple levels of abstraction, among which the representation of the nodes as concepts of a dynamical legal system. The “conceptual hubs” and “semantic subsets” of the US SC Corpus [7], along with the “hidden communities” of the French environmental code [8], recall the specific ‘interface’ of this kind of research. It does not deal with information *as* reality but *on* reality, namely, semantic information which is true or false. This very difference sets the specific level of abstraction that should be always taken into account when discussing the exchange of knowledge and methodologies between different branches of AI & Law and legal theory.

REFERENCES

[1] L. Floridi. The Method of Levels of Abstraction. *Minds and Machines*, 18(3): 303-329 (2008).

[2] N. Guarino, P. Giaretta. Ontologies and Knowledge Bases: Towards a Terminological Classification. In: *Towards Very Large Knowledge Bases*, N. Mars (ed.). IOS Press, Amsterdam, 25-32 (1995)

[3] J. Breuker, P. Casanovas, M.C.A. Klein, E. Francesconi. The Flood, the Channels and the Dikes: Managing Legal Information in a Globalized and Digital World. In: *Law, Ontologies and the Semantic Web. Channelling the Legal Information Flood*. J. Breuker, P. Casanovas, M.C.A. Klein, E. Francesconi (eds). IOS Press, Amsterdam, 3-18 (2009).

[4] M.E.J. Newman. Power Laws, Pareto Distributions and Zipf’s Law. *Contemporary Physics*, 46: 323-351 (2006).

[5] D. Rafailidis, Y. Manolopoulos. The Power of Music: Searching for Power-Laws in Symbolic Music Data, <http://delab.csd.auth.gr/papers/PC108rm.pdf> (2008).

[6] D. H. Zanette. Zipf’s Law and the Creation of Musical Context, http://arxiv.org/PS_cache/cs/pdf/0406/0406015v1.pdf (2008).

[7] M.J. Bommarito, D. Katz, J. Zelnher. Law as a Seamless Web? Comparison of Various Network Representations of the United States Supreme Court Corpus (1791-2005), <http://portal.aem.org/citation.cfm?id=1568234.1568270> (2009).

[8] R. Boulet, P. Mazzega, D. Bourcier. Network Analysis of the French Environmental Code. *AICOL-II*, Rotterdam, forthcoming.

[9] G. Ajani, P. Rossi. Sailing the Semantic Seas by Structural Vessels: Problems and Perspectives for Knowledge Discovery in the Legal Domain. *AICOL-II*, Rotterdam, forthcoming.

[10] U. Pagallo. As Law Goes by: Topology, Ontology, and Evolution. *AICOL-II*, Rotterdam, forthcoming.

[11] C. Anderson. The Long Tail: Why the Future of Business Is Selling Less of More. Hyperion, New York (2006).

[12] A.-L. Barabási. Linked: The New Science of Networks. Perseus, New York (2002).

[13] R.F. Voss, J. Clarke. /f Noise in Music and Speech. *Nature*, 149: 510-515 (1975).

[14] J.L. Hsu, C.C. Liu, A.L.P. Chen. Discovering Nontrivial Repeating Patterns in Music Data. *IEEE Transactions on Multimedia*, 3(3):311-325 (2001).

[15] H. Simon. The Architecture of Complexity. In: *Proceedings of the American Philosophical Society*, 106(6): 467-482 (1962).

[16] S. Chandler. The Network Structure of Supreme Court Jurisprudence. *The University of Houston Working Paper Series*, <http://ssrn.com> (2005).

[17] Th. Fowler, S. Jeon. The Authority of Supreme Court Precedent: A Network Analysis, <http://jh.fowler.ucdavis.edu> (2005).

[18] E. F. Keller. Revisiting ‘Scale-free’ Networks. *BioEssays*, 27: 1060-1068 (2005).

[19] S. A. Schettler. A Structured Overview of 50 Years of Small-World Research. *Social Networks*, 31: 165-178 (2009).

[20] U. Pagallo. “Small world” Paradigm and Empirical Research in Legal Ontologies: A Topological Approach. In: *The Multilanguage Complexity of European Law: Methodologies in Comparison*. G. Ajani, G. Peruginelli, G. Sartor, D. Tiscornia (eds.). European Press Academic Publishing, Florence: 195-210 (2007).

[21] S. Chandler. The Network Structure of the Uniform Commercial Code: It’s a Small World After All, <http://library.wolfram.com/infocenter/Conferences/5800/uccanalysis.pdf> (2005).

[22] Th. A. Smith. The Web of Law. *San Diego Legal Studies Research*, paper n° 06-11 (2005).

[23] R. Boyd, P. J. Richerson. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago (1985).

[24] J.S. Coleman. *Foundations of Social Theory*, Harvard University Press, Cambridge, Mass. (1991).

[25] S. Pinker. *The Stuff of Thought: Language as a Window into Human Nature*. Viking, New York, NY (2007).

Temporal Reasoning and MAS

Clara Smith¹, Antonino Rotolo², and Giovanni Sartor³

Abstract. In this paper we investigate if it is possible and useful to reason about time within social/normative multi-agent systems (MAS) by taking into account the general guidelines of tense logic. We focus on the combination of special-purpose logics: we provide a formal account in which a minimal temporalization helps in reasoning about time in an abstract way. We also explore a new variant of deontic tense logic by using a hybrid tense logic. The accounts provided allow to model temporal provisions within both particular norms and general legal principles, and also help in the detection of breaches of good faith and confidence.

1 INTRODUCTION

Both law and technology deal with temporal specifications. Relevant to the law are, among others, the prescription of actions to be accomplished within certain time, obligations with deadlines, the duration of the legal year, the duration of a leasehold in land, or an insurance policy during a specified period. For computing systems in general, finiteness and infiniteness of processes are of main interest, as well as the duration and successful termination of different kinds of atomic or complex transactions, the activity or inactivity of sessions, logins, the possibility of distributed or parallel processing, etc.

The relevance of the relation between time and deontic logic was well-settled in the seventies and eighties. Nowadays, several social and normative MAS capture diverse aspects of time mainly from a modelling point of view. However, consistent working applications of modal tense logics to those systems are still overlooked. Also, technical issues such as decidability of temporalized MAS are usually difficult to face.

For instance, the framework in [1] helps to model social constraints, roles, and states using concepts from legal and social systems, using the Event Calculus [2], a first-order formalism for reasoning about events. Time is explicitly codified inside constraints as in e.g. *HoldsAt*(permitted(*agent*, *action*),*t*) which generically specifies a permission: at time *t* it is true that the *action* is permitted for the *agent*. Given a temporal ordering of events we can find out the set of powers, permissions, obligations and sanctions each agent has at each point in time. As a second example, consider the account in [3]. It comprises a defeasible theory based on [4] which allows expressing some aspects of FIPA semantics for Agent Communication Languages (ACL). Instantants are modeled as timestamps labelling literals (a literal is an atomic proposition or its negation), e.g. *l:t* is meant to stand for ‘literal *l* holds at time *t*’.

Within both accounts, thinking back and forth along the time line may be a demanding enterprise. For example, moving along the flow of time means searching the knowledge base for rules and literals tagged with a given instant *t*, and performing the plausible deductions; then repeat with instant *t+1* (or resp. *t-1*) and so on (in line, see explanations to Example 3.3 in [4]). Typical temporal formulas such as: ‘it has always been the case that *p*’ or ‘it is always going to be the case that *p*’ need back and forward searches throughout the knowledge base (see the reading of rule *r*₁₀ in Example 3.3 in [4]).

Other MAS such as [5,6] handle in different ways some forms of temporal reasoning. The account in [5] for example, uses for its definition of trust some well-known concepts from dynamic logic (which can be seen as including temporal aspects.) In the survey in [7], its authors refer to various works incorporating aspects of time to overcome some inherent difficulties with standard deontic logic. For instance, the dyadic temporal operator proposed by J. A. van Eck [8] relativizes deontic statements to points in time where they should hold. The authors also refer to [9], which reduces deontic specifications to first-order temporal ones by interpreting the obligation to do *p* as the property that *p* must occur sometime in the future.

Theoretically speaking, the general idea of reasoning about time should extend any MAS consistently. From a computational standpoint, we believe that a minimal functionality for the automatic manipulation of time –i.e. from a basic modal perspective- should involve few technical adjustments of the MAS’s underlying logic. As an added challenge, we also argue that a temporalization straightforwardly helps in the detection of breaches of good faith and confidence in MAS.

We organize the work as follows. Section 2 describes two combinations of tense logics with multi-modal multi-agent logics. Section 2.1 uses a basic tense logic (understood in the classical way introduced by Prior [10]) to temporalize—in the sense given in [11, Section 2]—multi-modal MAS. This allows us to express abstract time provisions and lawful principles straightforwardly. In Section 2.2 we propose a new variant of deontic tense logic by considering an extension to modal logic called *hybrid* logic which allows us to name points in time. This way we gain a “HoldsAt” functionality. Section 2.3 extends both accounts in 2.1 and 2.2 by adding agent’s goals and beliefs. In all subsections we work with examples referring to Argentinean and to Italian Law, and also with a more general example from the literature. Examples are presented throughout the paper in an increasing level of complexity. In Section 3 we mention some decidability issues regarding the logics used throughout. Some conclusions end the paper.

2 TEMPORAL SPECIFICATIONS AND NORMS

46 Many MAS are designed as multi-modal systems [5,12,13,14]. Different modal operators define different agents’ features such as

¹ Faculty of Informatics and Faculty of Law, University of La Plata, Argentina. Partially supported by PICT 2006-00842 ANPCyT.

² CIRSIFID, University of Bologna, Italy.

³ European University Institute, Florence, Italy, and CIRSIFID, University of Bologna, Italy.

intentions, beliefs, agency, etc. Embeddings of tense logic into the basic modal language were first studied by S. K. Thomason in the mid '70s [15]; the approach has gained popularity recently, see [11,16,17,18]. It is pointed out in [7] that Thomason has argued that deontic logic requires a foundation in temporal logic, reducing the obligation of p to a temporal statement that p holds in all future worlds that would be reachable.

The description of a plain, basic logic of time, restricted to a traditional Kripke-style modal perspective, and which comprises a great variety of systems [17] is usually as follows: the basic temporal language is built using two unary operators, F and P , a set of propositional letters $p, q, r, \dots$, and Boolean connectives [18]. The intended interpretation of a formula Fp is 'p will be true at some Future time', and Pp is meant to stand for 'p was true at some Past time'. Duals for F and P are, respectively, G and H ('it is always Going to be the case' and 'it Has always been the case'.) This is the core language underlying the branch of modal logic called tense or temporal logic. The usual mathematical structures where temporal formulas are interpreted are bidirectional frames [18, pp. 21]. For now, let us assume a frame is a structure $\mathcal{T}=(T, <)$ where T is a set of instants of time and $<$ is a precedence relation such that if $s < t$ ($s, t \in T$) then we say that s is earlier than t [17]. The minimal pointwise tense logic is $K4$, which is complete w.r.t. the class of transitive frames. As one of our motivations is decidability, i.e. to provide decidable temporalized MAS, we will explore two possible methods for embedding temporal reasoning in multi-modal settings:

- In a first possibility, we add temporal modalities to a multi-modal language for agents already including operators for deontic provisions, agency, and, possibly, agents' goals and beliefs. Since this extension does not allow for explicitly referring to time instants in the language, it looks useful for representing abstract temporal properties or constraints, or for capturing uncertain time provisions such as regulations or obligations conditioned to undetermined events.
- A second possibility is to combine a multi-modal logic as mentioned above with a hybrid temporal logic (see, e.g. [18, pp. 434].) This logic explicitly identifies points in time and accounts for the moments at which an event happens, or the temporal scope of lawful provisions begins (or ends.)

2.1 Temporal Social/Normative MAS

The literature provides a number of different techniques for combining logics, such as products, fibring, fusion, and temporalization. Indeed, the reasoning patterns we present simply require the addition of a temporal logic on top of the logics underlying the MAS.

On account of [12]'s advice, let us work with a multi-modal approach for dealing with agents' attitudes. Assume we deal with a finite set of agents $A = \{x, y, z, \dots\}$ and a countable set P of atomic propositional sentences denoted by $p, q, r, \dots$. Complex expressions are formed syntactically from these, plus the following unary modalities we describe next.

A deontic operator O represents generic (legal/lawful) obligations, meaning "it is obligatory that" [7,12]. An operator for agency $Does_x \mathcal{A}$ represents successful agency i.e. agent x indeed brings about $\mathcal{A}$ [19]. For simplicity, we assume that in expressions like $Does_x \mathcal{A}$, $\mathcal{A}$ denotes behavioural actions concerning only single conducts of agents such as withdrawal, inform, purchase, payment, etc. (no modalized formulas occur in the scope of $Does$.) O is taken to be a classical KD operator. The logic of $Does$ is non-normal, closed under logical equivalence and amounts to the following schemata [20]: $Does_x \mathcal{A} \rightarrow \mathcal{A}$, $(Does_x \mathcal{A} \wedge Does_x \mathcal{B}) \rightarrow Does_x(\mathcal{A} \wedge \mathcal{B})$, and $\neg Does_x T$.

We now bring in the temporal modalities for writing temporal formulas; for example, the modality F mentioned above will allow us to characterise future obligations, as well as the future extinction of an obligation. Before we go to the semantic details, we present two examples.

Example 1. Obligations in the future. According to art. 566 of the Argentinean Civil Code, and according to art. 1183 of the Italian CC, a deadline may be established for complying with an obligation. Similarly, the emergence of obligations (and of the corresponding rights) can be postponed to a future moment in time; or a future moment may be established for extinction of obligations. Intuitively, time works in such scenarios as a modality *of the obligation*; the obligation is what is being modalized.

The temporal operator F allows us to express that an obligation will hold sometimes in the future: $F(O Does_x \mathcal{A})$ means that "it will be true in some future time that there is the obligation for agent x to do $\mathcal{A}$ ". It also allows us to express that there will be a future time in which an obligation will not hold: $F(\neg(O Does_x \mathcal{A}))$ means that "it will be true in some future time that there is no obligation for agent x to do $\mathcal{A}$ ". It may also apply to conditional obligations: $F(p \rightarrow O(Does_x \mathcal{A}))$ means that there will be an instant in the future when if p holds then there is the obligation for agent x to do $\mathcal{A}$. Similarly, $F(p \rightarrow \neg(O Does_x \mathcal{A}))$ means that there will be an instant in the future when if p holds then there is no obligation for agent x to do $\mathcal{A}$. As examples of this kind of normative statements consider the following:

- 1) $F(O(Does_j \text{ Pay}))$, there is a future moment in time in which agent j will have the obligation to pay;
- 2) $F(\neg(O(Does_j \text{ Pay})))$, there is a future moment in time in which there is no obligation for j to pay;
- 3) $F(Does_k \text{ Ask} \rightarrow (O Does_j \text{ Pay}))$, in the future there will be a moment in which, if k asks for the payment then j has the obligation to pay;
- 4) $F(Does_j \text{ Pays} \wedge (\neg(O(Does_j \text{ Pay}))))$, in the future there will be a moment in which it will be that case that j pays but he has no obligation to pay.

While such statements express interesting normative positions for the involved agents, they fail to capture temporally limited obligations, since they do not provide sufficient information to j and k . To be able to comply with an obligation starting in the future (1), j needs to know what is the precise time when the obligation will start to hold (it may not be enough to know that it

will hold sometime in the future.) Similarly, for being able to reject future compliance with an obligation terminating in the future (as in 2), and in, e.g. ‘I will not work for you tomorrow since my work contract expires today’) j needs to know when the obligation will terminate. For the obligation in (3) to be satisfied, it is sufficient that there is one future instant in which k ask for payment, or the payment is provided. On the contrary, usually, a conditional obligation to pay usually requires that payment is provided when one is asked for the first time (the creditor would not accept the reply: ‘ok, in the future there will be an occasion when I will pay when asked, but this is not that time; try again, you may be luckier!’).

These representations fail to capture the fact that an obligation usually does not hold for an instant, but it spans over a time interval, and similarly, it does not hold forever. A way to capture the persistence of an obligation (or its absence) consists in using G , the dual of F (Section 2.) The expression $G\mathcal{A}$ stands for “ $\mathcal{A}$ will always hold in the future”, e.g. $\text{Dies}_x \rightarrow G(\neg O(\text{Does}_x \mathcal{A}))$ means that if x dies, s/he will never have the obligation to do $\mathcal{A}$. For example, the death of a person who has been instituted as head of a life-rent extinguishes forever the creditor’s obligation to pay (art. 2070 ACC, art. 1873 ICC.) Simple formulas like this one are indeed powerful; this one can even be considered as a possible formulation for the legal institution called *prescription*, which in one of its forms establishes that “by the time designed by law, the debtor is free from its obligation” meaning that there will be a future moment in time in which the creditor cannot pursue her/his legal right in court (art. 4017 ACC, art. 2934 ICC.) However, to model a prescription more is required: we need to be able to represent the time point where the prescription operates, so that the obligation terminates. As we shall see in the following, for addressing the limitations just illustrated we supplement our logic with an explicit representation of time-points and intervals.

2.1.1 Semantics for Temporalized MAS

Regarding the semantics of the combination of logics we propose, we follow the *temporalization* technique [11,21] which, intuitively, amounts to place the temporal machinery on top of the MAS.

We consider both general and relativised obligations (O^a is a relativised deontic operator meaning “it is obligatory in the interest of agent a that” [22]). We restrict ourselves to well-formed atomic temporal formulas whose outermost symbol is a temporal operator. Formally, the behaviour of such a system is captured by a model $(T, <, g, t_0)$. The outer frame $(T, <)$ corresponds to the temporal evolution of the system; $t_0 \in T$ is the initial point in time. The system evolves through time in the sense that new generic/individual obligations/permissions are settled while some others become obsolete or prescribe. Following, g is a function such that for every $t \in T$, $g(t)$ returns a multi-relational model for the MAS itself. According to [14,20] we build a multi-relational frame of the form:

$$F = \langle A, W, O, \{O^i\}_{i \in G}, \{\mathcal{D}_i\}_{i \in G} \rangle$$

where:

- A is a finite set of agents;
- W is a set of situations, or possible worlds;
- O is the accessibility relation for the deontic operator, which is serial (standard KD semantics);
- $\{O^i\}_{i \in G}$ is the set of accessibility relations w.r.t. relativised obligations, which are serial (usual KD_n semantics);
- $\{\mathcal{D}_i\}_{i \in G}$ is a family of sets of accessibility relations D_i w.r.t. Does; which are reflexive, serial, and pointwise closed under intersection.

A multi-relational model is, in its turn, a structure of the form:

$$M = \langle F, V \rangle$$

where:

- F is a multi-relational frame as above, and
- V is a valuation function defined as follows:
 1. standard Boolean conditions;
 2. $V(w, O A) = 1$ iff $\forall v$ (if wOv then $V(v, A) = 1$);
 3. $V(w, O^i A) = 1$ iff $\forall v$ (if $wO^i v$ then $V(v, A) = 1$);
 4. $V(w, \text{Does}_i A) = 1$ iff $\exists D_i \in \mathcal{D}_i$ such that $\forall v$ ($wD_i v$ iff $V(v, A) = 1$);

This way, according to the notation in [12], $(T, <, g, t_0)$ amounts to a model for the logic of a temporalized MAS. Notice that although in the resulting logic we can write temporal formulas, the formulas of the nested MAS (which have no temporal operators) are unambiguously evaluated w.r.t. a given model $g(t)$.

2.2 Temporal Reasoning and the ‘HoldsAt’ Functionality

Terms can be certain or uncertain. This classification takes into account the moment at which the key event happens. If such moment is known by the time of arrangement (for example, at contract formation) the term is said to be certain, otherwise it is uncertain: although it will necessary happen, we do not know when it will (for example, a death.)

So far, we have been working with points in time throughout Section 2.1. However, we are not able to clearly identify such points in the account. This feature may appear to be a drawback for applications which consider selected time points; it is sometimes important to reason about what is going on at particular moments. Effective for representing provisions such as those in Section 2.1, basic temporal logic is plain and straightforward; it does not provide an analogous tool as the *HoldsAt*($\mathcal{A}, i$) predicate in [1] or the *!t* tagged literal in [3]. As Blackburn and Franceschet point out [11,18], hybrid logics –an extension of modal logics– helps us to deal with particular moments.

We will be working next with a *hybrid temporal logic*, which is an extension of the basic temporal logic used in Section 2.1. This logic will subsequently be the one used to temporalize. A *hybrid temporal language* helps us to treat points in time as “first class objects”, by naming them individually and directly. For example, if we are working with the domain of months, we can assign the

names $j, f, m, a, \dots$ to particular worlds in the domain of months the same way we give constant names to refer to specific elements in any domain (as we do with ‘0’ and ‘1’ within the integers.) Being able to locate points in time leads us to ask whether a certain event will happen in a given future moment, like in: “will we be having a rainy month of *April*?”, or if it has happened in a past time, as in: “were the company flights rescheduled in *August*?”. Finally, we may want to give a name to the current moment because, e.g., it is the moment in which we are in a position to fulfil our and other parties’ expectations; assuming a domain of days we can propose: “let’s sign the contract *today*, *December 27th*”, or: “you may stay and start working from *now*, *December 27th*, on”. This possibility looks necessary in order to be able to indicate a date in lawful acts.

For dealing with particular instants of time in our account, we proceed, technically, as follows. We take the basic modal language in Section 2.1 and add a second sort of atomic formulas. These atoms are called nominals, written $i, j, k, \dots$. A nominal names a world, or point, by being true in that point and nowhere else; so if i is a nominal the formula i holds if and only if the current location is named i . Old and new atoms are combined to form complex formulas in the usual way (as in $i \wedge p$.) For direct access to worlds, basic hybrid logics provides a ‘@’ operator. This operator, called a satisfaction operator, allows us to build formulas such as $@_i \mathcal{A}$, which is true at any point in a model if and only if $\mathcal{A}$ is satisfied at the unique point named by i . Thus we may eventually have a family of $@_i$ operators in our language; hybrid logics are essentially multi-modal logics. Satisfaction operators are normal modal operators. Note that $@_i \mathcal{A}$ expressions play the same role of $\text{HoldsAt}(\mathcal{A}, i)$ propositions and $i:t$ facts. Hybrid logics may also include the downarrow binder $\downarrow i$, which creates a name i and assigns it to the currently evaluated world. Two examples follow.

Example 2. Hybrid-temporalised MAS; termed obligations, and claimable obligations. Suppose we consider the domain of days. Then the formulas $\downarrow i.e \rightarrow @_i G(O(\text{Does}_x \mathcal{A}))$ and $\downarrow i.F(\downarrow j.e) \wedge @_i.G(j \rightarrow G(\neg(O \text{Does}_x \mathcal{A})))$ stand for “when the event e occurs, from that day on it will be the case that agent x is obliged to do $\mathcal{A}$ ”, and “from now on, there will be a day in which the event e will happen, and from that day on it will be the case that there is no obligation for x to do $\mathcal{A}$ ”, respectively. The fact that an obligation $\mathcal{A}$ is claimable by time i can be written as $@_i O \mathcal{A}$.

Example 3. Hybrid-temporalised MAS with relativised obligations; payment made before due time (Art 571 ACC.) The sentence

$$(\downarrow i.(\text{Does}_x \text{Pay}) \wedge @_i F(\downarrow j.O^x(\text{Does}_x \text{Pay}))) \rightarrow \\ @_i \neg G(O^x \text{Does}_x \text{Reimburse})^1$$

establishes that if agent x pays on day i and there is a future day in which it is obligatory in the interest of agent y that x pays (O^y)

¹ We may have also formalized this consequent as “there is no obligation in the interest of agent x for y to reimburse not in time i , neither after i and before j ”, meaning that the payment must have been done by time j .

expresses a relativised obligation, namely, the obligation meant to satisfy the interest of agent y , see [22]), then there is no obligation for y to reimburse such payment. It is plain to see that by using a hybrid logic we keep the modal temporal reasoning facilities and add a “HoldsAt” functionality.

As another example, consider the scenario where agent x works for agent y , starting on day i . This can be represented as $@_i G(O(\text{Does}_x \text{Workfor}(y)))$. Similarly, the fact that x will not be obliged to work for y from day j on can be written as $@_j \neg G(O(\text{Does}_x \text{Workfor}(y)))$. Note that though both prescriptions make sense separately, they lead to a contradictory situation: it cannot be the case that in the future there are instants in which x is obliged to work and obliged not to work for y . To avoid such type of contradictions, and to be able to expand our framework with obligations spanning until an event terminates them, we have to provide a further extension to our model. We will make use of the *until* modal operator.

Further extension. So far, we have been working with abstract temporal properties, and with uncertain time provisions. Then we added the possibility of explicitly referring to points in time. But for several MAS applications all these appear to be not enough. For example, we may want to be able to formally write sentences like “if agent a damages agent b , until the obligation is extinguished it is obligatory that a pays b ”, which responds to the more general temporal pattern “ p will be the case, and until that happens, q will hold”: $U(p, q)$. Those p properties are called *guarantee properties* in the computational literature [18]. While U searches forwards along a linear pointwise flow of time, the binary operator *since*: $S(p, q)$ is used to look backwards.

Let us see how to fit these operators in our hybrid-temporal multi-modal account. The satisfaction definition for U is:

$$t \models U(\phi, \psi) \text{ iff there is a } v > t \text{ such that } v \models \phi \text{ and for all } s \\ \text{with } t < s < v: s \models \psi$$

which it is to be interpreted on frames with structure $(T, <)$ as stated in the Introduction. Therefore, the above semantic definition for U clearly suits the temporal intended behaviour of the model $(T, <, g, t_0)$ we gave in Section 2.1.1. (Correspondingly, define S as a search in the opposite direction in the time line.)

It is well known that U and S are not definable in the basic modal language (see e.g. [17,18]). Moreover, proposition 7.10 in [18] proves that U is not definable over the real numbers with $<$, using F and P . On the other hand, F and P are definable in a language with U and S , thus such a language is stronger than the basic temporal language. Therefore, let us extend our language with the binary operators U and S . U and S -formulas are built straightforwardly from Boolean connectives and propositions in the usual way.

Example 4. Hybrid-temporal MAS with Until; Deal example. Agents a and b make the following deal: “ a has the obligation to pay b , until a does pay or June arrives”. Formally, this deal can be expressed as $U(((\text{Does}_a \text{Pay}) \vee \text{June}), O^b(\text{Does}_a \text{Pay}))$; recall that June is here a nominal (and nominals play the role of propositional constants.) A proper payment before the

forthcoming month of June will release agent a from his obligation. It follows from the example (and from U 's satisfaction definition) that the obligation to pay holds in all months previous to June and up to the month a pays (if on time); when June comes, the obligation will no longer hold.

2.3 Temporal Reasoning: MAS with Beliefs and Goals

We next test with a couple of examples our intuitions regarding the usefulness of the reasoning about time within MAS for detecting breaches of good faith and confidence. Objective good faith is usually described as correctness, i.e., as a standard of behaviour which is to be recognised not only with reference to legal rules, but also with reference to social norms or background rules (namely, to norms and rules that are not explicitly stated within a legal code) [23,24,25,26,27]. Article 1175 of the Italian Civil Code states that one should follow the *rules of correctness*. These require that the trustee takes care to avoid damaging the trustor, i.e., the trustee should take into account the legitimate interest and expectations of the other party.

For working with the next example we extend the language with an operator for goals: $\text{Goal}_x \mathcal{A}$. It aims to stand for “agent x has goal $\mathcal{A}$ ”, where $\mathcal{A}$ is a proposition.

Example 5. Temporalised MAS with Goal operators (in [28], revisited.) Agent a is debtor of agent b . At the time of prescription of b 's action against a , b asks a to pay his debt, making a know he will be sued otherwise. Agent a says that he has no money at the time but that he will have enough in six months, which largely exceeds the prescription limit. b accepts to informally relax the deal. Six months later, a refuses to pay b . Then b sues a .

Discussion. At time of prescription the following facts hold:

- $\text{O}^b(\text{Does}_a \text{ Pay})$, it is obligatory in the interest of agent b that a pays his debt;
- $\text{Goal}_a(\neg \text{Pay})$, agent a has as (inner) goal not to pay;
- $\text{O}^b(\text{Does}_a \text{ Pay}) \rightarrow \text{O}^b(\text{Goal}_a(\text{Does}_a \text{ Pay}))$, when there is an obligation, then the obliged agent has the good-faith obligation to have the goal to accomplish (correctness rule);
- $\text{G}(\neg \text{O}^b(\text{Does}_a \text{ Pay}))$, from now on it will always be the case that there is no obligation of a to pay, as the obligation prescribes;
- $\text{G}(\text{Goal}_a(\neg \text{Pay}))$, it will always be the case that a has as goal not to pay.

From the two first facts we get that a 's goal of not paying contradicts his obligation to pay. This is an indicative of a breach of a 's good faith by the time of prescription of b 's action. Six months later it holds that $\neg(\text{Does}_a \text{ Pay})$, which conforms the last fact in the list above and also contradicts the deal.

Following, let us consider what happens to the correctness rule $\text{O}^b(\text{Does}_a \text{ Pay}) \rightarrow \text{O}^b(\text{Goal}_a(\text{Does}_a \text{ Pay}))$: by not having the goal of paying at the time he had to pay, a has violated the correctness obligation. However, when the main obligation expires also the

correctness obligation does not apply any longer. Although there are indicatives of a 's not acting in good faith, b would lose the case. A lawyer would indeed say that, from the moment the action prescribes, the obligation becomes a *natural obligation* (see e.g. Art. 515 ACC.)—this means that its fulfilment is founded now not in law, but in a principle of equity (this means that a is not obliged to pay, but if he does then he has no right to have his money back.)

Some important aspects of good faith and trust may be analysed on the basis of the concept of *subjective good faith*, which requires the trustor's sincere belief that he is doing things without damaging others' rights. For discussing this richer scenario, extend again our language with the doxastic modality $\text{Bel}_x \mathcal{A}$, which is meant to represent that x has the belief that $\mathcal{A}$ ³. Let us now see an example illustrating how to combine hybrid temporal logic with agent's goals and beliefs.

Example 6. Hybrid-temporalised MAS with Until, Goal, and Bel operators (in [28, once again, revisited].) Agent a is debtor of agent b . By t_1 , the time of prescription of b 's action against a , b asks a to pay his debt, making a know he will be sued otherwise. Agent a says that he has no money now but that he will have enough in six months, by t_6 . Agent b accepts to relax the deal. By t_6 , a refuses to pay b . Then b sues a .

Discussion. The following hold:

- $\text{U}(\text{Expires}(\text{O}^b(\text{Does}_a \text{ Pay})), \text{O}^b(\text{Does}_a \text{ Pay}))$, it is obligatory in the interest of agent b that a pays his debt, until this obligation expires;
- $\text{Expires}(\text{O}^b(\text{Does}_a \text{ Pay})) \rightarrow \text{G}(\neg \text{O}^b(\text{Does}_a \text{ Pay}))$, when the obligation expires it does not hold any longer;
- $\text{O}^b(\text{Does}_a \text{ Pay}) \rightarrow \text{O}^b(\text{Goal}_a(\text{Does}_a \text{ Pay}))$, the aforementioned correctness rule;
- $@_{t_1} \text{Goal}_a(\neg \text{Pay})$, agent a has as (inner) goal not to pay;
- $@_{t_1} \text{G}(\text{Goal}_a(\neg \text{Pay}))$, it will always be the case that a has as goal not to pay;
- $@_{t_6} \text{Expires}(\text{O}^b(\text{Does}_a \text{ Pay}))$; the deal expires at t_6 .

By time t_6 agent a has achieved his goals: he will no longer be obliged to pay ($\text{G}(\neg \text{O}^b(\text{Does}_a \text{ Pay}))$), and he will never have to pay, because the obligation has expired (i.e. the deal is over.)

Let us assume on the contrary that b sues a , at a time t_0 before the expiration time t_1 . In this case, it holds that:

- $@_{t_0} \text{O}^b(\text{Does}_a \text{ Pay})$;
- $@_{t_0} \neg \text{Does}_a \text{ Pay}$;
- $@_{t_0} \text{Goal}_a(\neg \text{Does}_a \text{ Pay})$;
- $@_{t_0} \text{O}^b(\text{Goal}_a(\text{Does}_a \text{ Pay}))$.

Consequently, at t_0 , agent a violates both the obligation to pay and the good-faith correctness rule to have the goal to pay. Note here that the corresponding sanction can be even more severe that it would be if he failed to accomplish his obligation in good faith, i.e., without having the goal to do so.

² As classically established, Goal_i is a K_a operator. The multi-relational frame already defined should correspondingly be extended with $\{G_i\}_{i \in G}$, a set of accessibility relations w.r.t. goals.

³ Respectively, extend again the multi-relational frame, now with a set of accessibility relations w.r.t. Bel , which are transitive, Euclidean and serial.

3 DECIDABILITY ISSUES

Regarding issues of decidability w.r.t. to our proposals in this work, it is well known that the logic that results from the combination through temporalization of a decidable logic and a basic tense logic is also decidable. PSPACE algorithms have been devised for a number of well-known logics including the temporal counterparts of K, T, K4 and S4. Theorem 7.1 in [18, pp. 436] settles that the satisfiability problem for basic hybrid logics is PSPACE-complete. Finite model checking algorithms for temporalizations –in the significance given in this paper- are available in [11]. Moreover, that paper suggests a temporalization using a hybrid logic. Furthermore, [29] provides terminating tableau systems for a number on non-transitive hybrid logics extending K such as the logic of irreflexive, antisymmetric frames; also [29] provides (for hybrid tense logic enriched with a universal modality) a terminating tableau calculus for the logic of transitive frames. Finally, U and S-formulas are complete w.r.t. frames $(T, <)$ such that $(T, <)$ is a well-ordered flow of time called Dedekind complete order, such as the total order of the natural numbers [11]. We consider all this background provides a strong preliminary platform where to build proof procedures for temporalized MAS such as those in this paper.

4 CONCLUDING REMARKS

In this paper we addressed some forms of temporal reasoning within MAS, and suggested a perspective on how these combinations can be used for modelling aspects of time within lawful provisions, obligations, legal principles, and good faith. The main collaboration is the new variant of deontic tense logic using hybrid logics. The temporalizations we suggested provide evidence that minimal adjustments are required for existing frameworks such as those in, e.g., [14] to deal with both a minimal temporal reasoning functionality and a ‘HoldsAt’ device. We applied a methodology which falls under the general name of combination of logics (see [11,30]): we combined special-purpose logics of restricted expressive power.

The systems we obtain are simple from the logical point of view; their simplicity is support for their usefulness and robustness, and also keeps the systems manageable, decidable, and suitable for further studies and extensions. The use of U leads us to consider an even more prominent role for time periods within normative MAS; we have treated intervals not as primitive objects but built them on top of a more basic element, the point in time. We did not include complex definitions for concepts such as trust, agreement, commitment, roles, or speech acts, which directly increase the complexity of the MAS. These further extensions, i.e. how to put to work all (or some of) these concepts together, are matter of our future investigations.

REFERENCES

[1] A. Artikis, J. Pitt, M. Sergot. Animated Specifications of Computational Societies. AAMAS 2002, ACM.
[2] R. Kowalski, M. Sergot. A logic-based calculus of events. New Generation computing, 4(1):67-96, 1986.

[3] G. Boella, G. Governatori, J. Hulstijn, R. Riveret, A. Rotolo, L. van der Torre. Time and Defeasibility in FIPA ACL Semantics. WLIAMAS 08.
[4] G. Governatori, G. Sartor, A. Rotolo. Temporalised normative positions in defeasible logic. Proceedings of ICAIL 2005, ACM Press.
[5] C. Castelfranchi, R. Falcone, ‘Social trust: a cognitive approach’, in *Trust and Deception in Virtual Societies*, Springer, 2001.
[6] R. Demolombe, C-J. Liao. A Logic of Graded Trust and Belief Fusion. 4th Wshop on Deception, Fraud and Trust in Agent Societies. 2001.
[7] J.-J. CH. Meyer, R. J. Wieringa. Deontic logic: a Concise Overview. In: *Deontic Logic in Computer Science: Normative System Specification*. J.-J. CH. Meyer, R. J. Wieringa (eds.) 1993, Wiley.
[8] J. A. van Eck. A system of temporally relative modal and deontic predicate logic and its philosophical applications. *Logique et Analyse*, 100, 1982, 249-381.
[9] T. Maibaum, J. Fiadeiro, and S. Khosla. Temporal reasoning over deontic specifications. In *Deontic logic in computer science: normative system specification*, John Wiley and Sons Ltd., Chichester, 1994.
[10] A. N. Prior. Time and Modality. Oxford Press, 1957.
[11] M. Franceschet, A. Montanari, M. De Rijke. Model checking for combined logics with an application to mobile systems. *Autom. Softw. Eng.* 11(3), 289-321, 2004.
[12] A.J.I. Jones, M. Sergot. A Logical Framework. Open Agent Societies: Normative Specifications in Multi-Agent Systems, Wiley, 2003.
[13] A. Rotolo, G. Sartor, C. Smith. Good Faith in Contract Negotiation and Performance. *IJBPM Journal*, Inderscience. 4(3) :154-173. doi: 10.1504/IJBPM.2009.030983.
[14] A. Rotolo, C. Smith. Collective Trust and Normative Agents. *Logic Journal of IGPL* 2010; doi: 10.1093/jigpal/jzp076.
[15] S. K. Thomason. Semantic Analysis of Tense Logics. *JSL*, 35:150-158, 1972.
[16] R. Goldblatt. Mathematical Modal Logic: A View of its Evolution. Handbook of the History of logic. V. 6. 2005, Elsevier.
[17] Y. Venema. Temporal Logic. L. Goble ed., The Blackwell Guide to Philosophical Logic. Blackwell Publishers, USA. 203-223, 2001.
[18] P. Blackburn, M de Rijke, Y. Venema. Modal Logic. Cambridge, 2001.
[19] D. Elgesem. The Modal Logic of Agency. *Nordic Journal of Philosophical Logic*, 1997.
[20] G. Governatori, A. Rotolo. On the axiomatisation of Elgesem’s logic of agency and ability. *Journal Philosophical Logic*, 2005, 34: 403-431.
[21] M. Finger, D. Gabbay. Adding a temporal dimension to a logic system. *Journal of Logic, Language and Information*, 2:203-233, 1992.
[22] H. Herrestad, C. Krogh. Deontic Logic relativised to Bearers and Counterparties. J. Bing and O. Torvund eds., *Anniversary Antology in Computers and Law –TANO 1995*, pp 453-522.
[23] D. Memmo, G. Sartor, G. Quadri. ‘Trust, reliance, good faith and the law’, in *Trust Management*, Springer. 2003.
[24] H. MacQueen. ‘Good faith in the Scots law of contract: an undisclosed principle?’. In *Good Faith in Contract and Property Law*, Hart, 1999.
[25] R. Brownsword, N. Bird, G. Howells. Good faith in contract: concept and context, in *Good Faith in Contract*, Ashgate. 1998.
[26] M. Storme. Good faith in the contents of contracts in European Private Law. *Elec. Journ. Comparative Law*, 2003.
[27] A. Musy. The good faith principle in Contract Law and the precontractual duty to disclose, in *Global Jurist Advances*. 2001.
[28] S. Whittaker, R. Zimmermann. Good Faith in European Contract Law. Cambridge University Press, 2000.
[29] T. Bolander, P. Blackburn. Terminating Tableau Calculi for Hybrid Logics extending K. www.elsevier.nl/locate/entcs. 2007.
[30] C. Areces, C. Monz, H de Nivelle, M. de Rijke. The Guarded Fragment. Ins and Outs. In J. Gerbrandy, et al eds., *JFAK. Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*, Vossiuspers, AUP, Amsterdam, 1999.

The fifth NorMAS workshop shows the breadth of the research area, ranging from virtual organizations, formal models and simulation studies to gaming agents and social psychology.

In “Contract argumentation in virtual organizations”, Criado, Heras, Argente, and Julian propose a novel dialogue game protocol combined with argumentation schemes for contract agreement within dynamic organizations controlled by norms to enable these virtual organizations to evolve over time using norm adaptation.

Burgemeestre, Hulstijn and Tan address a similar issue, that of open norms. In “Norm emergence in regulatory compliance” they discuss how norms can be discussed and interpreted, resulting in a norm emergence process. A case study of adoption of Authorized Economic Operator legislation in the EU customs domain is presented resulting in two observations: (1) norm adoption is guided in a top-down fashion, by EU guidelines and auditing standards, and (2) consensus about what counts as acceptable is reached by interaction among companies, consultants, and auditors. To explain the norm emergence process the authors state that two kinds of theories need be combined: (i) theories about the specification, prescription and implementation of norms (one-to-many, top-down), and (ii) theories about norm adoption, consensus reaching and norm distribution (many-to-many, bottom-up).

In “Non-taking up as a normative behavior”, Saillard uses normative multiagent systems to describe and analyze why individuals abstain from exercising their (social) rights. Components of the analysis include the autonomy of the agents, their cognitive activity and the progressive construction of effective social rights.

Mahmoud, Griffiths, Keppens and Luck present an empirical analysis of Axelrod’s model of norms and meta-norms which enables norm establishment in populations of self-interested individuals. In their paper “An analysis of norm emergence in Axelrod’s model” they explore some of the key assumptions made in previous evaluations of the model. These include the dynamics of norm emergence and the occurrence of norm collapse when applying the model over extended durations. They also investigate in detail the reasons for norm collapse in extended applications of the model to show that both the level of mutation in the population and the precise nature of the reproduction mechanism are significant.

In “Modelling the social fabric for normative NPCs in MMOGs”, Johansson and Verhagen sketch the requirements for believable agents or “NPCs” in multiplayer online games. Current agents lack flexibility and learning on the social level and the authors state that to develop the next generation of “smart” NPCs assistance is needed from the NorMAS community to provide formal models of and reasoning mechanisms for normative agents in MMOGs, combined with theories and models from the social sciences.

Finally, invited speaker Corinna Elsenbroich presents “The Uniqueness of normative behaviour”. Observing that normative behaviour has been one of the earliest topics investigated by agent-based models, she finds that many of the very early models are extremely simple and is amazed how much the aggregate behaviour of these simple models resembles actual normative behaviour outcomes.

But many factors unique to normative behaviour have not been captured yet and by presenting normative behaviour from a social psychology view she offers a solution to part of these missing issues. An integration of some cognitive capacities, such as reasoning about values, would bring NorMAS research a long way towards capturing the unique features of normative behaviour.

Harko Verhagen

Paul Davidsson

Davide Grossi

Contract Argumentation in Virtual Organizations

N.Criado, S.Heras, E.Argente, V.Julián¹

Abstract. Virtual Organizations should evolve along time in order to face with unforeseen changing situations. This evolution implies the adaptation of norms, so agents must be able to consider dynamic norms in their decisions. With this aim, this paper proposes a new dialogue game protocol together with argumentation schemes for contract agreement inside dynamic organizations controlled by norms.

1 Introduction

Virtual Organizations (VOs) have been employed as an abstraction for modelling open agent societies. They include the integration of both the organizational and individual perspectives and also the dynamic adaptation of models to organizational and environmental changes [4]. In this way, it is necessary to detect situations of interest and to manage them maximising the adaptation flexibility and capacity. The organization adaptation affects to, among other aspects, its norms, agreements, commitments and topologic structure.

Taking into account the challenges and difficulties arisen by the dynamic adaptation of VOs, our proposal consists of a new dialogue game protocol and the adaptation of previous argumentation schemes in order to allow agents to reach agreements considering the existence of dynamic norms which control their behaviours.

This paper is structured as follows: Section 2 provides a theoretical definition of VO together with a model of norms for controlling agents belonging to VOs. Section 3 describes the proposed argumentation schemes for contract agreement. This argumentation protocol has been applied into a case-study in Section 4. Finally, a discussion of related proposals and future works are shown.

2 Virtual Organization Model

A VO is formally defined in [3] as follows:

Definition 1 (Virtual Organization). *A Virtual Organization (VO) is defined as a tuple $VO = \langle Ag, Ro, Roles, \leq_{prior}, OU, N, Se, ER, Act, Cap, Des, Prob \rangle$ where:*

- *Ag is the set of agents which are the members of the VO.*
- *Ro is the set of roles which have been defined in the VO.*
- *Roles : $Ag \rightarrow 2^R$ is a function that matches an agent with the set of roles which it is currently playing.*

- *$\leq_{prior}$ is a total order relationship which expresses priority among roles.*
- *$OU = \{ou_0, ou_1, \dots\}$ is the set of organizational units which have been defined in the VO, where each $ou_i \subseteq (Ag \cup OU)$.*
- *N is the normative context of the VO, i.e. the set of norms that control the operations carried out in the VO.*
- *S is the set of services defined inside the VO.*
- *ER is the set of resources located in the VO environment.*
- *Act is a set of first order atomic formulae which represent primary individual actions. These actions include the request and provision of services and the perception and access to the environmental resources.*
- *Cap : $Ag \rightarrow 2^{Act}$ is a function which maps each agent with actions it is able to perform.*
- *Des : $Act \rightarrow [-1, 1]$ is a function that assigns a real value for representing desirability/undesirability of actions.*
- *Prob : $Act \rightarrow [0, 1]$ is a function that assigns a probability to an action occurrence.*

Due to space reasons, other features of the VO model, such as dynamicity, have not been described here. However, this theoretical definition of a VO allows defining: i) the *Structural* dimension ; ii) the *Functional* dimension ; iii) the *Environmental* dimension ; and (iv) The *Normative Context* (N). Next, a complete definition of the normative context is provided.

2.1 Normative Model

In this work norms are classified into two main levels according to the norm scope:

Definition 2 (Normative Context). *The set of norms of a VO is defined as $N = IN \cup CN$ where:*

- *IN are the institutional norms which have been explicitly promulgated by the institution; their violation is considered as an illicit act and it might lead to institutional sanctions.*
- *CN are norms derived as a consequence of an interaction among agents*

Definition 3 (Institutional Norm). *An institutional norm ($n_i \in IN$) is defined as $n_i = \langle D, C, T, S, E \rangle$ where:*

- *$D \in \{O, F\}$; being O and F deontic operators for representing obligations (O) and prohibitions (F), respectively.*
- *$C \in Act$ is a formula that represents the action that must be carried out in case of obligations; or that has to be avoided in case of prohibitions.*
- *$T \in Ro$ is the target of the norm; i.e. the role affected by the norm.*

¹ Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, Camino de Vera s/n. 46022 Valencia (España), email{ncriado,sheras,eargente,vinglada}@dsic.upv.es

- $S \in Act$ is an expression which describes the action that will be carried out as sanction in case of norm violation.
- $E \in Act$ represents the norm expiration action.

For simplicity, norms in our model have not an activation condition. Thus, they are always active until the expiration action holds. Moreover, norms have been defined as a deontic control over a single action. However, this work can be extended for considering a more complex notion of norms.

Semantics of an institutional norm depends on the concrete deontic type. For example semantics of an institutional obligation is defined as follows:

$$\langle O, C, T, S, E \rangle \equiv \forall a \in Ag : T \in Roles(a) :$$

$$(Do_a(C) \rightarrow \mathcal{F}) \mathcal{U} (E \wedge (\neg \mathcal{F} \rightarrow Do_{Institution}(S)))$$

$\mathcal{U}$ is the temporal binary operator *Until* and $\mathcal{F}$ is the norm fulfilment proposition. $Do_a(C)$ is the predicate for representing an agent a carrying out action C . Therefore, after the promulgation of the norm, the performance of the norm condition by an agent which is affected by the norm ($a \in Ag : T \in Roles(a)$) implies the fulfilment of the norm ($Do_a(C) \rightarrow \mathcal{F}$). Once the norm is deactivated (E), sanction is performed by the institution if the norm has not been fulfilled ($\neg \mathcal{F} \rightarrow Do_{Institution}(S)$).

Contract Norms (CN) are normative prescriptions that emerge as a result of the interaction among agents. Therefore, this level of norms is formed by the contracts reached by agents along their execution.

Definition 4 (Contract). A contract ($n_c \in CN$) is defined as $n_c = \langle D, C, a, b, S, R, E \rangle$ where:

- $D \in \{O, F\}$ represents the deontic type of the contract.
- $C \in Act$ is the previously defined norm condition.
- $a, b \in Ag$ are the contract addressee and beneficiary agents, respectively.
- $S, R \in Act$ are expressions which describe the actions that will be carried out in case of contract violation or fulfilment, respectively.
- $E \in Act$ represents the norm expiration action.

Semantics of contracts also depends on the concrete deontic type. For example, semantics of a prohibition contract is defined as follows:

$$\langle F, C, a, b, S, R, E \rangle \equiv$$

$$(Do_a(C) \rightarrow \mathcal{V}) \mathcal{U} (E \wedge (\neg \mathcal{V} \rightarrow Do_b(R)) \wedge (\mathcal{V} \rightarrow Do_b(S)))$$

$\mathcal{V}$ is the violation proposition. Thus, after the contract creation, occurrence of the condition implies a violation ($Do_a(C) \rightarrow \mathcal{V}$). Once the contract expires, violations will be sanctioned and contract fulfilment will be rewarded ($E \wedge (\neg \mathcal{V} \rightarrow Do_b(R)) \wedge (\mathcal{V} \rightarrow Do_b(S))$).

2.2 Norm Conflicts

Our proposal takes as a reference the work on norm conflict detection and resolution described in [9]. In this work they define substitutions as:

Definition 5 (Substitution). A substitution σ is a finite and possibly empty set of pairs x/y where x is a variable and y is a term.

The application of a substitution is defined as:

1. $c \cdot \sigma = c$ if c is a constant.
2. $x \cdot \sigma = y \cdot \sigma$ if $x/y \in \sigma$; otherwise $x \cdot \sigma = x$.
3. $p^n(\rho_0, \dots, \rho_n) \cdot \sigma = p^n(\rho_0 \cdot \sigma, \dots, \rho_n \cdot \sigma)$, where p^n is a n -ary predicate symbol.

Then the unification relationship is defined as follows:

Definition 6 (Unification). The $unify(C, C')$ relationship holds iff there is a possibly empty substitution σ such that $C \cdot \sigma = C' \cdot \sigma$.

Taking these previous definitions as a reference, we define the notion of norm conflict in our model. In our proposal a norm conflict can be an *Institutional Conflict* or a *Contract Conflict*. Next, each type of norm conflict is formally defined.

Definition 7 (Institutional Conflict). Any contract $n_c = \langle D, C, a, b, S, R, E \rangle$ is in conflict ($IC(n_c, n'_i)$) with an institutional norm $n'_i = \langle D', C', T', S', E' \rangle$ if: i) they define opposite deontic constraints $D \neq D'$; ii) over the same action $unify(C, C')$; iii) the contract is addressed to an agent which plays the norm target role ($T' \in Roles(a)$); and iv) the institutional norm is active ($\neg E'$).

Definition 8 (Contract Conflict). A contract $n_c = \langle D, C, a, b, S, R, E \rangle$ is in conflict ($CC(n_c, n'_c)$) with a previous contract $n'_c = \langle D', C', a', b', S', R', E' \rangle$ if: i) they define opposite deontic constraints $D \neq D'$; ii) over the same action $unify(C, C')$; iii) both contracts are addressed to the same agent $a = a'$; and iv) the previous contract is active ($\neg E'$).

3 Argumentation For Contract Agreement

Our main objective in this research is to solve a problem of contract agreement. In our initial settings we have two agents $i, j \in Ag$ that want to reach an agreement over the acceptance of a particular contract. Therefore, the *proponent* agent i has to persuade the *opponent* agent j to accept a contract $n_c \in CN : n_c = \langle D, C, j, i, S, R, E \rangle$. Thus, the proponent agent i starts a dialogue with the opponent agent j and proposes it norm n_c . In this section, we show the type of arguments and attacks to these arguments that agents can interchange during the dialogue in the context of our normative model. Then, we propose a dialogue game protocol that controls the agents' interaction. The study of the internal preferences of agents, their reputation and the trust that they have in other agents are out of the scope of this work. Here we have focused on defining the type of arguments and attacks that agents can put forward and the rules of the dialogue that manage the interchange of these arguments and attacks in the normative model proposed.

3.1 Initial Arguments

Following our normative model, the opponent agent j could have several possible types of arguments to reject the acceptance of norm n_c :

- A1: There is an institutional norm $n'_i \in IN, n'_i = \langle D', C, T', S', E' \rangle$ which is in conflict with n_c (i.e. $IC(n_c, n'_i)$).

- A2: There is a previous contract agreement by which agent j accepted a contract $n'_c \in CN$, $n'_c = \langle D', C', j, k, S', R', E' \rangle$ which is in conflict with n_c (i.e. $CC(n_c, n'_c)$).
- A3: There is a lack of interest in the reward awarded for the acceptance of this norm (i.e. $Des(R) == 0$), but the opponent specifies in what case it would be interested in the reward (R') so as to accept n_c .
- A4: There is a lack of interest in the reward awarded for the acceptance of this norm (i.e. $Des(R) == 0$).
- A5: Agent j cannot accept norm n'_c since it has no capabilities to fulfil this norm (i.e. $C \notin Cap(j)$).

Some of these arguments fit stereotyped patterns of human reasoning that argumentation theory has modelled as *Argumentation Schemes* [10]. Concretely, arguments A1 and A2 fit the normative version of the *Argument from Oppositions* scheme. The scheme has the following structure:

Major Premise: If X receives the evaluation W' , X cannot receive the opposite evaluation W at the same time and in the same respect.
Minor Premise: X receives the evaluation W' .
Conclusion: X cannot receive the opposite evaluation W at the same time and in the same respect.

Then, if we adapt this scheme to our settings, we have that:

Scheme 1. "If the opponent agent j accepts the norm n' , with $n' \in \{n'_c, n'_i\}$, agent j cannot accept the conflicting norm n_c at the same time and at the same respect".

Thus, if agent j receives n_c and at the same time has an institutional norm n'_i or a previous contract n'_c that conflicts with n_c , agent j can conclude that it cannot accept n_c .

If the opponent agent rejects norm n_c due to argument A3, it initially rejects the norm but provides the proponent with an alternative. This alternative would consist of a condition that, if fulfilled by the proponent, the opponent would accept norm n_c . In this case, the argument fits the structure of the *Two Person Practical Reasoning* scheme:

Major Premise: X intends to realize A , and tells Y this.
Minor Premise: As Y sees the situation, B is a necessary (sufficient) condition for carrying out A , and Y tells X this.
Conclusion: Therefore, X should carry out B , unless he has better reasons not to.

Adapting this scheme to our settings, it is reformulated as:

Scheme 2. "Agent i intends to realize contract n_c and tells agent j this. As j sees the situation, offering a new reward R' is a necessary condition for carrying out contract n_c , and j tells i this. Therefore, i should offer a new reward R' , unless it has better reasons not to".

Thus, with this argument the opponent agent points out the proponent a possible way of persuading it and carrying on with the dialogue. Also, this scheme has a set of *Critical Questions* (CQ) associated [10]. These questions are stereotyped ways of attacking the conclusion of the argument and they will be analysed in the next section.

Finally, by default we assume that the preferences and capabilities of an agent do not change during the dialogue (but they can change when the dialogue finishes). Therefore, A4 and A5 are indisputable arguments that end the dialogue.

3.2 Attacks

Depending on the type of argument that the opponent agent has for rejecting the acceptance of contract n_c , the proponent agent still has the opportunity of trying to persuade it to accept it. This is the case of arguments of types A1, A2 and A3. On the opposite, if the opponent did not provide the proponent with any alternative way of accepting norm n_c in A5 or it rejected the norm by putting forward A4, these arguments are considered indisputable and the dialogue will end here. Thus, arguments A1, A2 and A3 admit replies (i.e. attacks) from the proponent.

In the case of A1 and A2, the *Argument from Oppositions* scheme does not have associated critical questions. However, it can be still attacked if the proponent agent thinks that there is an exception to its conclusion. This exception poses a new argument for not paying attention to the conflict between norms and the proponent could send it to the opponent, trying again to persuade it to accept norm n_c . This pattern of reasoning fits the *Argument For an Exceptional Case*:

Major Premise: If the case of x is an exception, then the established rule can be waived in the case of x .
Minor Premise: The case of a is an exception.
Conclusion: Therefore, the established rule can be waived in the case of a .

Adapting this scheme to our settings, we have that:

Scheme 3. "If the case of A1 (A2) is an exception, then the conclusion of the *Argument From Oppositions* scheme can be waived in the case of A1 (A2). Thus, if we have that A1 (A2) are exceptions to the established rule, then the conclusion of the *Argument From Oppositions* scheme can be waived".

In the context of our normative model, possible exceptions to A1 are:

- E1:** The desirability of reward for fulfilling the contract n_c is greater than the undesirability of the sanction applied for the non-fulfilment of the institutional norm n'_i (i.e. $Des(R) + Des(S') > 0$).
- E2:** Actually, the sanction associated for the non-fulfilment of the conflicting institutional norm n'_i is never applied (i.e. $Prob(S) = 0$).

Also, possible exceptions to A2 are:

- E3:** The reward for fulfilling the contract n_c is greater than the sanction and the lost of reward implied by non-fulfilment of the contract n'_c (i.e. $Des(R) > Res(R') - Des(S')$).
- E4:** The role r_i of the proponent agent i has more priority than the role r_k of the agent k that agreed the previous contract (and its associated norm n'_c) with the opponent agent (i.e. $\exists r_i \in Roles(i) : \forall r_k \in Roles(k) : r_i \geq_{prior} r_k$).

If the proponent agent attacks the argument A1 or A2 of the opponent agent by putting forward the conclusion of the *Argument for an Exceptional Case*, the opponent can surrender to the attack and accept the norm n_c or else it can reject again the norm. A possible way of rejecting this attack could be to pose one of the critical questions of the scheme. If we adapt these questions to our normative model we have

the following possible defeats²:

CQ1: Can evidence that the conclusion of the *Argument From Oppositions* scheme does not apply to this case be given?

We have modelled possible attacks to the possible exceptions by adapting the type of attack defined in the CQ1. Thus, the opponent can attack exceptions by requesting the proponent to show evidence for the exception. Then, even if the proponent shows evidence for the exceptions, the opponent could still reject the norm. Possible reasons could be, for instance, if it does not agree with (or believe) the evidence provided by the proponent, if it prefers observing the institutional norms, no matter how many times sanctions are applied or what the reward for fulfilling contracts is or if it has another argument of the type explained in the previous section to reject again the norm.

Finally, the proponent agent could attack argument A3. In this argument, the opponent is willing to accept norm n_c with the condition that the proponent offers a new reward R' , following the conclusion of the *Two Person Practical Reasoning* scheme. We assume that offering a new reward R' is the only condition that the opponent puts for accepting contract n_c . Thus, if the proponent agent decides not to offer a new reward, it just rejects A3 without giving alternatives for dispute. However, if it finally agrees to change the reward of the norm, it proposes again the new norm $n'_c \in IN, n'_c = \langle D, C, j, b, S, R', E \rangle$ to the opponent.

3.3 Dialogue Game Protocol for Contract Agreement

Following the approach of [8], we have modelled the agents' dialogue by means of a formal dialogue game for contract agreement that allows agents to reason about argumentation schemes. The protocol assumes the existence of a proponent agent i that tries to agree a contract $n_c \in CN$ with an opponent agent j . The interaction protocol between the agents of the network has been modelled with the components for dialogue games identified by McBurney and Parsons in [5]. The rules of the protocol are founded in the logic patterns of reasoning that agents could follow in a situation like this.

- **Commencement Rules:** The dialogue starts when the proponent agent i sends to the opponent agent j a proposal for the acceptance of a new contract n_c . Then, j receives the proposal and has to make a decision on whether accepting the norm or rejecting it.
- **Locutions:**
 - **Statements:** they are the permissible locutions in the dialogue and their compounds (*propose(norm)*, *accept*, *reject* and *assert(argument)*). Arguments can be arguments of the type described in section 3.1, information about norms or evidences of the type shown in section 3.2, depending on the stage of the dialogue when this locution is stated.
 - **Withdrawals:** *noCommit()* is the locution that agents use when they cannot answer the locution that they have

received. In the case of the proponent, it withdraws its proposal and ends the dialogue. In the case of the opponent, it withdraws its last argument for rejecting the norm and either rejects it again (due to a different reason), accepts it or ends the dialogue. Also, this locution is used when the opponent does not want (or cannot) provide an answer to the proponent, but still does not accept the contract norm.

- **Questions:** the locution *assert(norm)?* asks the opponent agent for information about the previous contract that conflicts with the proposed norm. Agents are not committed to answer questions. Thus, the opponent agent is not committed to give any information about its norms to the proponent.
- **Attacks:** the locution *pose(exception)* poses an attack to the argument of the type A1 and A2.
- **Challenges:** *why(locution)?*, where locution can be a rejection or an exception attack, requests arguments that support them.
- **Commitment Rules:**
 - The assertion of the locution *propose(norm)* commits the proponent and the opponent to observe the rules of the dialogue game.
 - The assertion of the locution *noCommit()* by the proponent, withdraws its proposal and ends the dialogue.
 - The assertion of the locution *noCommit()* by the opponent, withdraws its last argument for rejecting the norm (if any).
 - The assertion of the locution *why(locution)?* commits the recipient either to provide an argument for the rejection or the attack or else, to withdraw it.
- **Dialogue Rules:** The rules of the game define which are the allowed locutions in each step of the dialogue. Figure 1 shows the sequence diagram that provides a graphical representation of the dialogue. In this figure, objects represent steps and arrows represent transitions between steps, which are labelled with the associated locution.
- R1:** The proponent agent opens the contract agreement process proposing the acceptance of the contract norm n_c to the opponent (*propose(n_c)*).
- R2:** The opponent agent can *accept* the contract and end the dialogue, end the dialogue without providing an answer (*noCommit()*) or reject the contract norm (*reject*).
- R3:** The proponent can either withdraw its proposal and end the dialogue (*noCommit()*) or ask the opponent for providing an argument for the rejection of the contract (*why(reject)?*).
- R4:** The opponent agent can either withdraw the rejection and come back to R2 (*noCommit()*) or *assert* an argument (i.e. A1, A2, A3, A4 or A5) that supports the rejection of the contract norm.
- R5:** The proponent agent can either withdraw its proposal (*noCommit()*) and end the dialogue or *pose* an attack with exceptions E1 or E2.
- R6:** The proponent agent can either withdraw its proposal (*noCommit()*) and end the dialogue or it can

² The original definition of the critical questions is available at [10]

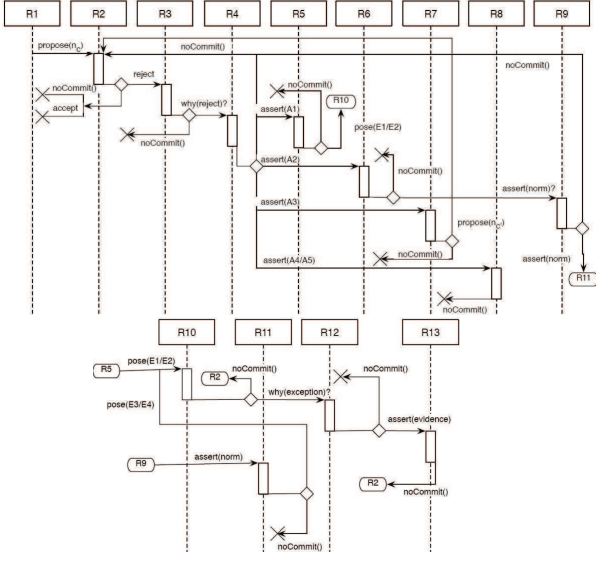


Figure 1. Sequence Diagram of the Interaction Protocol

ask the opponent for more information about its contract n'_c that conflicts with the proposed contract norm. ($assert(norm)?$) By default, the proponent does not have any information about the sanction of this norm or the role of the agent that enacted it, which are necessary information to pose exceptions E3 or E4.

- R7:** The proponent agent can either withdraw its proposal ($noCommit()$) and end the dialogue or make a new proposal n'_c with the new reward R' ($propose(n'_c)$).
 - R8:** The proponent agent has to withdraw its proposal and end the dialogue, since the opponent does not give any chance for dispute.
 - R9:** The opponent agent can either withdraw the rejection and come back to R2 or provide the proponent with information about its previous contract that conflicts with the proposed contract norm ($assert(norm)$).
 - R10:** The opponent agent can either withdraw the rejection and come back to R2 or ask the proponent for providing an argument for the exception attack received ($why(exception)?$).
 - R11:** The proponent agent can either withdraw its proposal and end the dialogue or pose an attack with exceptions E3 or E4, coming back to R10.
 - R12:** The proponent agent can either withdraw its proposal and end the dialogue or provide the opponent with an evidence that supports the exception attack that it posed ($assert(evidences)$).
 - R13:** The opponent agent withdraws its rejection and comes back to R2.
- **Termination Rules:** The dialogue ends either when the opponent agent accepts the contract norm proposed by the

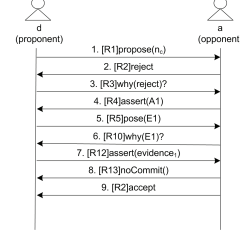


Figure 2. Case-Study Dialogue

proponent agent or when the proponent agent withdraws its proposal. In addition, both agents can end the dialogue in any moment without reaching an agreement about the contract by putting forward a $noCommit()$ locution. However, the opponent agent also uses this locution to withdraw the last argument that it provided for rejecting the norm and reject it with any other argument. To avoid infinite dialogues, the opponent agent is not allowed to use the same argument for rejecting the contract twice. Also, after a maximum time is exceeded, the dialogue ends without reaching to a contract agreement.

This proposal has been applied into a case study in the following section.

4 Case Study: the m-Water Domain

The management of water in a basin is a challenge of significant social relevance which can be seen as a VO. In this sense, *irrigators* which belong to the same area of a basin can be organized forming VOs or *irrigator communities*. These communities act on behalf of their members by defending their rights and interests. However, each community can impose some norms or restrictions to their members. This case study illustrates how an *irrigator* agent which belongs to one *irrigation community* reaches an agreement on a contract with other agents, considering the existence of both norms imposed by the community and previous contracts.

Usually, irrigator communities forbid farmers to sell their production on their own, since they obtain better sale conditions if they sell their vegetables as a single entity (i.e. they do not compete among them):

$$n_i = \langle F, sell(Product, Buyer), Farmer, \\ pay(community, 1000), - \rangle$$

Therefore, if a farmer is caught selling his products on his own, then the irrigator community will punish him by paying 1000 euros ($pay(community, 1000)$).

In this situation, let us suppose that an agent (a) which is a farmer belonging to the irrigator community ($farmer \in Roles(a)$) wants to reach a contract with another agent (d) which is interested on buying vegetables.

According to the dialogue game protocol, the proponent agent (d) sends a contract proposition to the farmer agent a (Figure 2 Scenario 1 Message 1). Thus, n_c is a contract for

selling vegetables:

$$n_c = \langle O, \text{sell}(\text{vegetables}, d), a, d, -, \text{pay}(a, 5000), \rangle$$

According to this contract, if the farmer fulfils the contract, it would be rewarded by the distributor. More concretely, agent d would pay 5000 euros to the farmer if it sells its collected vegetables. Since agent a does not want to violate norms imposed by its community, then a rejects the contract (Scenario 1 Message 2). Then the proponent agent asks for more information about the rejection of the contract (Scenario 1 Message 3). In this case, the opponent agent shows argument $A1$ that claims that there is an institutional norm which forbids the contract conditions. More concretely, there is an institutional conflict ($IC(n_c, n'_i)$) since there is a substitution $\sigma = \{Product/vegetables, Buyer/d\}$ such us $unifies(\text{sell}(Product, Buyer), \text{sell}(vegetables, d))$. After observing the norm, the proponent poses and exception attack to this argument. Next, the opponent requests an argument for the exception attack (Scenario 1 Message 6). Then the proponent shows the evidence ($evidence_1$) that the sanction imposed by the community is lower to the contract reward ($Des(\text{pay}(a, community, 1000)) < Des(\text{pay}(d, a, 5000))$). Finally, agent a withdraws its rejection and accepts the contract (Scenario 1 Messages 8 and 9).

5 Discussion

The work presented in this paper proposes a new dialogue game protocol together with argumentation schemes for contract agreement inside dynamic VOs controlled by norms. Also, we show how argumentation schemes can be used to model the agents' reasoning over the acceptance or rejection of norms. Due to space reasons, the previous section describes the usage of our framework in a simple case study. In a near future, we plan to apply this approach into different case studies for evaluating its performance in practical situations.

Argumentation theory has been applied to MAS in several works. In [1], the formal representation of the procedural aspects of an argumentation protocol was proposed. The protocol distinguishes between constitutive rules, which define the protocol itself and the physical environment where it is executed, and the normative context that could make protocol actions obligatory or permitted. The main difference with our work is the external perspective adopted in this protocol. While they only take into account externally observable behaviours, we focus on the internal reasons that could make an agent to accept or deny a norm associated to a contract agreement process in a virtual organization. In addition, their protocol specifies actions that can be performed over propositions whereas our model determines the type of arguments that can be used to reject or accept norms.

In [6] a framework for conflict resolution in normative systems is presented. In this work, authors also adopt an externalist view, where agents present arguments to decide between executing an individual desired goal or observing an institutional norm. These norms are enforced through punishments and rewards modelled as the goals of enforcement agents that act as mediators in the dialogue. However, norms that emerge from interactions among agents are not considered.

On the area of norm negotiation, [2] presented a negotiation model for the acceptance of new social norms based on

what they call the social delegation cycle. This model focuses on the emergence of norms in normative MAS, from the initial negotiation of social goals, to the final acceptance of their associated social norms when they fulfil the desires of the agents. This is considered as a social process, where agents collaborate to decide social norms. However, our model focuses on the formalisation of the individual arguments that can be exchanged between agents to persuade an opponent to accept an already existent norm.

A work that also proposes argumentation schemes to reason about norms was proposed in [7]. Unlike our model, this work focuses on the internal reasoning that an agent can perform in view of an internal normative conflict, taking into account priorities over norms, and the effects of these norms. On the opposite, our model is intended for virtual organizations, where several agents can argue about the individual acceptance or rejection of norms that other agents propose.

Finally, our protocol takes into account the internal reasons that could make an agent accept or deny a contract norm, but it may be argued that it is not a realistic assumption that the opponent will reveal its reasons for rejecting a contract proposal. However, we make this assumption in the basis that the agents really want to reach an agreement. Therefore, this agreement could hardly be reached if the agents are not willing to reveal their internal reasons to accept or reject a contract. This is the most reasonable way of finding an alternative that could fit the interests of both proponent and opponent agents.

ACKNOWLEDGEMENTS

This work was partially supported by the Spanish government under grants CONSOLIDER-INGENIO 2010 CSD2007-00022 and TIN2009-13839-C03-01, by GVA funds under PROMETEO 2008/051 project and by the FPU grant AP-2007-01256 awarded to N. Criado.

REFERENCES

- [1] A. Artikis, M. Sergot, and J. Pitt. An executable specification of a formal argumentation protocol. *Artificial Intelligence*, 171(10–15):776–804, 2007.
- [2] G. Boella and L. van der Torre. Norm negotiation in multiagent systems. *Int. J. Cooperative Inf. Syst.*, 16(1):97–122, 2007.
- [3] N. Criado, E. Argente, V. Julian, and V. Botti. Designing virtual organizations. In *PAAMS*, volume 55 of *Advances in Soft Computing*, pages 440–449, 2009.
- [4] V. Dignum and F. Dignum. A landscape of agent systems for the real world. Technical report, Inst. Information and Computing Sciences, Utrecht University, 2006.
- [5] P. McBurney and S. Parsons. Dialogue games in multi-agent systems. *Informal Logic*, 22:2002, 2002.
- [6] S. Modgil and M. Luck. Argumentation based resolution of conflicts between desires and normative goals. In *ArgMAS 2008. Revised Selected and Invited Papers*, 2009.
- [7] N. Oren, M. Luck, and T. J. Norman. Argumentation for normative reasoning. In *BRMAS at AISB*, 2008.
- [8] C. Reed and D. Walton. Argumentation schemes in dialogue. In *Dissensus and the Search for Common Ground*, pages 1–11, 2007.
- [9] W. W. Vasconcelos, M. J. Kollingbaum, and T. J. Norman. Resolving conflict and inconsistency in norm-regulated virtual organizations. In *AAMAS*, pages 91–99, 2007.
- [10] D. Walton. *Appeal to Expert Opinion*. Penn State Press, 1997.

Norm Emergence in Regulatory Compliance

Brigitte Burgemeestre¹, Joris Hulstijn¹ and Yao-Hua Tan^{1,2}

Abstract.

A trend in regulatory compliance is to issue 'open norms': norms which leave room for contextual interpretation about their implementation. Subjects must demonstrate to the regulator how they have interpreted the norms and collect evidence to show that the implementation is operationally effective. Subjects communicate frequently about the interpretation of open norms and copy compliance behaviour of others. We view this social norm interpretation process as a form of *norm emergence*. We have performed a case study of adoption of AEO (Authorized Economic Operator) legislation in the EU customs domain. We make two observations: (1) norm adoption is guided in a top-down fashion, by EU guidelines and auditing standards, (2) consensus about what counts as acceptable is reached by interaction among companies, consultants, and auditors. In this paper we review of some literature about norm emergence in the multi agent systems field, which can explain these observations. It turns out that we need to combine two kinds of theories: (i) theories about the specification, prescription and implementation of norms (one-to-many, top-down), and (ii) theories about norm adoption, consensus reaching and norm distribution (many-to-many, bottom-up).

1 INTRODUCTION

A trend in regulatory compliance is to formulate legislation as open norms that leave room for contextual interpretation about their implementation. Open norms do not describe in detail what is permissible or prohibited; they indicate what behaviour is expected by abstract principles or general objectives. For example, a recipe may prescribe to add salt "according to taste". How much salt must be added depends on the cook and his guests. Through experience a cook learns what reasonable quantities of salt are appropriate for different kinds of dishes.

In this work, we discuss the emergence of norms in human practice. We consider two types of agents: agents that implement norms (subjects) and agents that audit norm compliance (regulators). Open norms assume that subjects are able to interpret and implement legislation appropriately [2][3]. Whereas rule-based norms are quite straightforward to implement and require little interpretation, open norms need to be tailored to an agent's specific situation. Furthermore, open norms need to be translated into concrete constraints or rules before they can be implemented in IT systems [11]. From the regulator's perspective open norms also require a different approach to norm enforcement. Compliance with rules can be simply checked; compliance with principles or general objectives requires interpretation of norms and an evaluation of

the implementation in a certain context. The regulator thus needs to get an insight in how and why a company has implemented principles or general objectives as certain rules in its business processes to determine compliance [3]. Even though the norms are adapted to the unique circumstances of an agent, norm compliance has to be enforced in a fair and consistent way. Regulators can therefore also learn from best practices developed in the field, and use them as benchmarks to evaluate future implementations of open norms.

When legislation is relatively new, both types of agents may lack experience and knowledge to work with the norms. Subjects need examples of appropriate norm implementations from regulators. Regulators on the other hand, need input from the field to determine fair evaluation criteria to assess norm implementation. Regulators are thus awaiting experiences from the field, and the agents are awaiting implementation guidance and criteria. To overcome this potential deadlock situation, we need to understand the emergence of norms to be able to accelerate both of these learning processes. Understanding these processes is relevant for research in social simulation [12] and also for the specification of normative multi-agent systems with certain desirable properties.

Current research in multi agent systems (MAS) addresses norm evolution and adoption either as norm emergence (bottom-up) or norm prescription (top-down)[12]. We argue that an integrated approach is needed to study norm emergence more realistically: bottom up processes are likely to be bounded by law(s) and top down processes may be influenced by bottom-up forces. Especially for compliance with open norms, where implementations are instantiations of open norms (top-down), but where consensus emerges about how little or how much compliance effort is considered acceptable (bottom-up), it is important to address both processes simultaneously.

In this paper report on an extensive case study about the adoption of new legislation regarding Authorized Economic Operators (AEO), see [2][3]. AEO legislation aims to secure the supply-chain while reducing the administrative burden for companies. Companies may apply for an AEO certificate, when they can demonstrate to be trustworthy to customs concerning safety and security in the supply chain. The EU requirements for an AEO certificate are relatively general and abstract, so they form a nice example of open norms. Companies must identify risks and determine control measures to comply with the legislation. Customs must evaluate whether the control measures are sufficient given the risks. In the case study, we study what norm emergence processes take place in the interaction between customs and companies. We identify which aspects of the case study are covered by the literature and could be used as a basis for an integrated model on norm emergence.

The remainder of the paper is organized as follows. First, we describe processes of norm emergence encountered in the case study. In Section 3 we discuss a selection of literature on norm emergence. Next, we discuss if the literature explains the findings of our case study and identify blank spots. Section 5 presents our conclusions and provides directions for future work.

¹Department of IT Audit, Faculty of Economics and Business Administration, Vrije Universiteit, Amsterdam, the Netherlands. Email: {cburgemeestre, jhulstijn}@feweb.vu.nl

²Department of Technology, Policy and Management, Delft University of Technology, the Netherlands. Email: y.tan@tudelft.nl

2 CASE STUDY: CUSTOMS REGULATION

The case study is a form of qualitative empirical research about the adoption and enforcement of legislation in practice, namely the Authorised Economic Operators (AEO) legislation [7, 8]. In previous work we have reported on the same case study [3]. Here we focus on the emergence of norms in the interaction among customs agencies (regulator) and companies (subjects), who are guided in some cases by advice from legal consultants. In particular, we study the emergence, standardization and spreading of various implementations of open norms among companies, as well as the consensus building about the acceptability of those implementations among customs' auditors in different regions and between auditors and companies.

2.1 Data Collection

Data for this case study was collected by the following methods: document analysis, semi-structured interviews and observation. We studied internal and public documents from the Dutch tax and customs administration (Dutch TCA) and a petrochemical company (PCC), which recently applied for the AEO certificate. We conducted five interviews with auditors of Dutch TCA about PCC's application, and joined Dutch TCA auditors on two visits to PCC. We also studied PCC's AEO application documents. Except for the visits, we tape-recorded all interviews with the participants' prior agreement. Minutes were made of meetings.

2.2 Norm emergence in the AEO case

Recently, new European customs legislation has been introduced to enhance the security of the international supply chain. The legislation introduces certificates for so called Authorized Economic Operators (AEO). The idea is that security can be increased through a combined effort of business and government. Companies that can be trusted, i.e. meet the requirements to receive an AEO certificate, benefit from reduced customs inspections, simplified procedures and therefore a reduced administrative burden. Customs can focus their attention on non-certified companies. At the European level guidelines have been defined that leave room for local implementations. National customs' agencies are free to develop their own control strategies within the European legislation [10]. Therefore these EU guidelines can be characterized as open norms. Based on the EU guidelines, customs specify evaluation criteria for audits; companies on the other hand interpret the guidelines as rules for implementations, and produce context dependent instantiations.

A company may qualify as AEO, when it can demonstrate to have (i) an appropriate record of compliance concerning tax and customs issues, (ii) a satisfactory system of managing commercial, and where appropriate, transport records, which allows appropriate customs controls, (iii) proven financial solvency and (iv) appropriate safety and security standards [7,8].

The implementation of the open AEO guidelines is a struggle for both customs and companies. Customs have to decide which implementations conform to the open norms and may serve as benchmarks for the field. Since the legislation is relatively new, there are no best practices or guidelines for implementation. This is true in particular for guideline (iv) about safety and security. The knowledge of customs auditors about this topic is often tacit and therefore difficult to share. Moreover, the implementation of

the AEO guidelines requires domain specific knowledge to translate the open norms into company or industry specific solutions. For example, security risks in the scrap metal industry are of a completely different nature than risks in the pharmaceutical industry and therefore require a different set of security measures. Customs are able to recognize most risks but often lack the knowledge to come up with company specific solutions. On the other hand, controllers and the internal audit department of companies have more knowledge about the business and are therefore more capable of doing a risk assessment. So expertise from both customs and companies is needed to instantiate the abstract AEO guidelines with specific procedures, information systems and security measures.

Whereas in some countries customs take the lead in specifying the open norms, Dutch customs aim to do this in cooperation with the business sector. Dutch customs does not prescribe detailed rules for companies but let them propose their own implementations of the open norms. A self-assessment, in which companies need to identify their current position and necessary actions to comply with the legislation, is part of the certification procedure to become AEO. In the Dutch practice, the self-assessment concerning guideline (iv), about safety and security, must contain a risk assessment performed by the company itself, to identify the particular risks for the company. Given those risks, the company must show which security measures they have chosen to mitigate those risks, and provide evidence that these measures are operationally effective.

One of the ideas underlying the control approach of Dutch customs is that companies already take actions to control their business processes, for example because of accounting rules or other compliance regulations about health or safety, and that these controls (partially) cover the requirements of the AEO legislation. These existing controls may or may not be accepted by Dutch customs as a sufficient instantiation of the AEO guidelines. Effort is reduced for customs as they get input from the field on contemporary control solutions. Companies, on the other hand, only have to adapt their systems on those points that do not yet conform to the AEO guidelines.

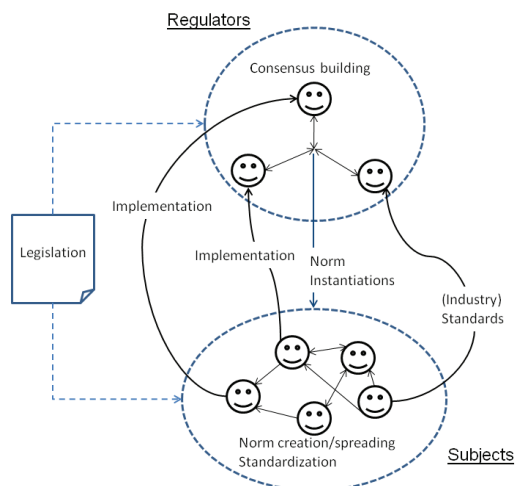


Figure 1. Levels involved in norm emergence

In the AEO case we see that customs and companies together create norms (Figure 1). Norms emerge when companies develop implementations, interact with other companies and agree on industry standards. These implementations and standards are then proposed to customs. Customs guide the norm emergence process by approving or disapproving the proposed implementations as instantiations of the open norms. Different regional offices of the customs organization also have to agree which implementations are acceptable. Approved implementations may function as benchmarks for future audits, or for future implementations of the guidelines [9].

Interviews with customs officials and companies revealed some concrete examples of norm emergence in the AEO case:

1. Implementation of appropriate safety and security measures is a requirement for the AEO certificate [7]. Examples of best practices in transport security are ISO/PAS:28000 and ISPS (International Ship and Port Facility Security Code). Companies often have limited or no security knowledge. Security officers from companies often rely on technical experts employed by vendors of security solutions, such as electronic key systems, RFID solutions and tracking and tracing capabilities. So vendors influence the emergence of new technologies and their adoption as best practices.
On the other hand we see that companies tend to imitate the behaviour of their competitors, and do not want to invest more money in security than necessary. Security is not a competitive advantage. Typically, in security *"a fence should not be lower than that of the nearest neighbours"*.
2. One of the criteria to become AEO is to have a satisfactory system of managing commercial and, where appropriate, transport records, which allows appropriate customs controls [7]. Customs procedures affect various departments of a company (financial, transport, operations etc.) and information is therefore often scattered over several business systems. Since customs operations are a small part of the business (PCC had only two customs employees) and outsourced companies are not willing to invest in specialized customs software packages, it is hard to collect the data. An intermediate solution to reduce the workload for companies is to provide customs with an overview of how the information is scattered over different business systems instead of creating a new system that contains all the information.
3. As we stated above, a risk assessment in which companies identify their exposure of potential violations has become an official part of the Dutch AEO certification procedure [8]. Companies report the findings to customs and have to explain how the measures they have taken address the risks.
4. Open norms may result in various interpretations. For example, AEO legislation requires that companies have appropriate measures in place to prevent "tampering with cargo units" (article 14 k) [7]. Companies can for example implement security seals on containers, or instead hire trusted transport companies, recruiting well-trained personnel. Customs assess whether these implementations comply with the norm. Companies in turn can demand explanation why a company is not AEO eligible. In case of a negative decision, a more elaborate explanation is needed for open norms than for rule-based norms, for which the decision is clear. Over time, public evidence about positive or negative decisions will to consensus about what is considered an appropriate implementation of an open norm, c.f. [9].

3 THEORIES OF NORM EMERGENCE

In the case study we observe that norms emerge through discussions about compliance between and among companies and regulators. This is a different process from the usual top-down norm specification and organizational implementation, as we find it in legal sources, work on compliance [11] and also in work on electronic institutions, e.g. [1][6][14]. Such frameworks assume that norms lead to the adoption certain goals, or control objectives, which are subsequently turned into constraints on actions or control measures. But here ideas on norm implementation can originate bottom-up from the subjects who like to comply. For example, in the case study we observe that companies copy security measures from competitors. On the other hand, the process is also different from the process of emergence of social norms or conventions, as we find it in work about social networks or emergence of social conventions [12]. Unlike such frameworks, which allow basically any behaviour to emerge as a norm, here the emergent norms are constrained by a related abstract norm specified by an institution. We therefore study some representative examples of different kinds of research in these areas, and formulate hypotheses to make the prediction of the theories more concrete. The hypotheses can then be compared to the findings in the case study.

3.1 Social mechanisms of norm emergence

Savarimuthu and Cranefield [12] provide an overview of recent work on distributed normative behaviour in simulation studies. The distributed approach focuses on the bottom-up emergence and spreading of norms in a system. For social norms they propose a norm-life cycle model that consists of four phases: (1) norm creation, (2) norm spreading, (3) norm enforcement and (4) norm emergence. They consider two variants of norm emergence: (i) the spreading of a norm in an agent society, until the norm is recognized and followed by most agents, and (ii) bottom-up creation of norms by cognitive agents. Such agents derive a 'proposed norm' based on their cognitive ability or come up with an alternative norm (creation phase) and then help in the emergence of that norm (emergence phase).

For each of the phases of the norm-life cycle model, they present an overview of the mechanisms which are found in norm simulation research. The mechanisms fall into the following categories: social power (punishment, leadership), learning (machine learning, imitation), reputation, off-line design, cognitive, emotion-based, dynamic and static social network typologies, and cultural and evolutionary aspects.

Hypothesis 1 (life-cycle): We expect to find behaviour corresponding to the phases (1) norm creation, (2) norm spreading, and (3) norm enforcement.

Hypothesis 2 (mechanisms): We expect to find norm emergence to be enhanced by mechanisms, such as social power, reputation and imitation, and we expect it to be dependent on the agent's position in the social network typology.

3.2 Emergence and immergence.

As has been mentioned in [12], actions of individual agents influence the evolution of norms in a society and on the other hand, norms in a society affect individual agents. Conte et al [4]

describe these processes and the micro- macro link, in more detail. They define emergence as the generation of macro social properties by micro-social behaviour. Individual agents intentionally or unintentionally produce effects by creating the conditions for them to arise. Conte et al. also describe the downward causation in which an emergent, macro-social property generates effects at the lower level agents. Down-ward causation occurs in two ways: by a simple and a complex feedback loop.

In the simple loop the emergent effect retroacts on the lower level by determining a new property of the generating system. A property could be for example increased negotiation power in a dependence network. In the complex loop, the emergent effect determines new properties by means of which the effect is reproduced again, which includes two sub processes: immergence and second order emergence.

"Immergence is the process by means of which the emergent effect modifies the way of functioning of the generating system, affecting its generating rules or mechanisms in such a way that it is likelier to be reproduced."

Second order emergence (or incorporation) is the process by means of which an emergent effect is recognised by the producing systems and by this means, the effect is likelier to be reproduced" [4].

Hypothesis 3 (Emergence and immergence): We expect to find (i) macro behaviour, induced by micro decisions, and (ii) adaptation of micro behaviour as a result of macro effects.

3.3 Instantiating and implementing open norms

The body of research on the instantiation and implementation of open norms in MAS is limited. Dignum [6] observes that norms in regulations are (on purpose) specified at a high abstraction level to account for many different situations which may occur over time. Regulations are specified on a higher abstraction level than the level on which the processes and structure of the institution are specified. Norms thus need to be translated and adapted to a certain domain before they can be implemented. The paper discusses different levels of abstraction. Institutions are considered to have a predefined objective and a set of values that direct towards fulfilling that objective. Attached to each value is a list of norms that contribute to that value. *"Norms contribute to a value if fulfilling the norm always leads to states in which the value is more fully accomplished than the states where the norm is not fulfilled"* [6]. The set of norms as a whole 'defines' the meaning of the value in the context of the institution. The behaviour of agents who operate in the institution must be in line with the objectives, values and norms defined by the institution. In order to check norm compliance and act on possible violations, the abstract norms have to be translated into concrete specifications of behaviour. *"Concrete norms pertain to actions that are described in terms of the ontology of the institution and from which therefore the meaning and effect is known or they pertain to situations that can be checked directly by the institute"* [6]. Based on a categorization of abstractness Dignum indicates how translations can be made between the types of abstract norms and more concrete norms. For the translation from concrete norms to the implementation level, Dignum suggest that for each concrete norm a rule is needed that either specifies part of the norm enforcement procedures of the institution, or specifies triggers that signal a violation of the norm and the expected reaction of the institution.

Aldewereld et al [1] describe the implementation of a norm enforcement mechanism for electronic institutions that is based on detecting violations of norms and reacting to these violations instead of restricting agent behaviour up beforehand. The paper proposes to use integrity constraints and dialogical constraints to implement such a mechanism. They point out that the implementation of norms requires an operational semantics. The declarative nature of norms is necessary for reasoning about what is considered legal or illegal, while the operational semantics defines how norms are to be implemented, i.e. what to do when norms are in fact violated. Norms that are described in abstract terms, first need to be 'contextualised' before they can be implemented. This contextualisation consists of two steps: (1) interpret the abstract concepts of the norm and link them to concrete concepts used in the institution, using counts-as operators, and (2) adding procedural information and artefacts to the institution to allow enforcement of the norm.

In [14] the operationalization of institutional norms in MAS is discussed in more detail. Norms are categorized depending on the actors involved, verifiability of states and actions in norm expressions, and temporal aspects. For each aspect guidelines on the implementation of norms are proposed.

Hypothesis 4 (Operationalization): Open norms are interpreted in two steps: (i) abstract concepts are instantiated by concrete objects (counts-as), and (ii) procedural information and enforcement artefacts are added to the institution.

3.4 Integrating social conventions and legal norms

Whereas both the prescriptive and conventional view on norms is researched, they are often studied independently of each other. Catselfranchi and Conte [5] try to establish connections between social conventions and prescriptions and model norms as cognitive objects. Their hypothesis is that the emergence of norms is intertwined with the emergence of normative beliefs. Unlike the conventionalists who see norm emergence as a process of agents imitating observable behaviour of other agents, they see emergence as a non-continuous phenomenon. A social norm is seen to imply a belief that a given behaviour is generally prescribed within an (agent) community. This behaviour is executed because and as long as it is believed to be obliged. The consequence of conforming to a given conduct – believed to be prescribed – promotes the act of prescribing it, thus contributing to its spreading.

Hypothesis 5 (Normative beliefs): Norm emergence co-occurs with the creation of a normative belief in a community, i.e. a convention that the norm is generally going to be followed.

4 RESULTS

In the case study we identified four concrete examples of norm emergence that take place in the adoption of European customs legislation about AEO. We observed: 1. The influence of vendors and imitation behaviour in security standards, 2. the development of individual solutions to comply with administrative requirements, 4. risk assessment as a means to justify chosen implementations to customs, and 5. discussions between customs and companies on the appropriateness of implementations of open norms.

We reviewed a selection of literature from the multi agent systems field to assess if current studies are able to explain the phenomena encountered in the case study. We grouped the literature into four topics: 1 Social mechanisms of norm emergence, 2 Emergence and immergeance, 3 Instantiation and implementation of open norms and 4 Integrating social conventions and norms. We formulated five hypotheses based on the literature. An overview of the case study findings with regard to the hypotheses is presented in Table 1 below.

Hypothesis	Source	Confirmed	Case study findings
1. Life-cycle	[12]	yes	- spreading: security standards are distributed by vendors or by imitation - creation: intermediate solution to enable customs' administrative controls - enforcement: customs (dis)approving norm implementations
2. Mechanisms	[12]	yes	- social power: customs (dis)approving norm implementations - reputation: hiring security specialists, relying on standards - imitation: fence as high as nearest neighbours - topology: applying industry specific solutions
3. Emergence and immergeance	[4]	yes	- approved solutions (jurisprudence) define boundaries for interpreting open AEO norms - AEO norms put requirements on administrative organization
4. Operationalization	[1] [6] [14]	yes	- Access controls are instantiations of "appropriate measures" - Security seals that prevent tampering are attached to containers
5. Normative beliefs	[5]	yes	- (Additional) measures need to be implemented by a company if a risk assessment indicates that they are not compliant with AEO legislation

Table 1: Overview of hypotheses and case study findings

We will now discuss the case study findings with regard to each of the hypotheses.

Hypothesis 1 (life-cycle): We expect to find behaviour corresponding to the phases (1) norm creation, (2) norm spreading, and (3) norm enforcement.

In the case study we identified behaviour corresponding to all three phases. An example of norm creation is the intermediate solution of enabling customs to perform administrative controls,

by collecting data from existing administrative software, instead of having to implement an integrated compliance system. Forms of norm spreading can be recognized in the distribution of security standards, by the influence of technology vendors (star topology) or by imitation (peer to peer). Norm enforcement by approving or disapproving certain solutions is the main task of Dutch customs. We can thus confirm Hypothesis 1. This entails that to build more realistic regulated MAS applications all phases should be taken into account.

Hypothesis 2 (mechanisms): We expect to find norm emergence to be enhanced by mechanisms, such as social power, reputation and imitation, and we expect it to be dependent on the agent's position in the social network typology.

We observed that customs, through their statutory power of approving or disapproving implementations can accelerate or slow down the emergence and spreading of certain solutions. Companies are more likely to choose approved implementations when they want to apply for AEO. They indicated in interviews to be interested in concrete advice from Dutch customs. Reputation-based mechanisms we encountered in the hiring of so security specialists, and the reliance on standards. Solutions proposed by well-known security vendors or standardization organizations are favoured over developing solutions internally. Mechanisms of imitation are found in the copying of security solutions of competitors, summarized by the quote about neighbours' fence height. The effect of the network topology can be found in branch organizations promoting industry specific solutions. A branch organization, a popular vendor, or a dominant company in the market will typically be found in the centre of a star topology. The different mechanisms defined in the hypothesis are thus observed. Insights from these studies can be used to identify agents in the environment that play a key role in (the acceleration of) norm emergence. An interesting finding is that the regulator can also be a part of a norm emergence mechanism, as enforcer but also as distributor.

Hypothesis 3 (Emergence and immergeance): We expect to find (i) macro behaviour, induced by micro decisions, and (ii) adaptation of micro behaviour as a result of macro effects.

Concerning (ii), our case study shows that the AEO legislation imposes clear requirements on the administrative organization of a firm, for instance considering reliability of commercial and transport records, enforced by an internal controls system, segregation of duties, information security measures etc. Concerning (i) we find that jurisprudence in the form of implementations that are known to be accepted or refused as proper instantiations of the AEO guidelines by the customs authorities, will aid in further defining the boundaries of the legislation. For example, if a risk assessment motivates that the risks for a certain industry are lower, requirements for security measures may decrease for that industry. Decisions by individual agents produce macro level effects by creating the conditions for them to arise or be generated [4]. Hypothesis 3 is confirmed.

Hypothesis 4 (Operationalization): Open norms are interpreted in two steps: (i) abstract concepts are instantiated by concrete objects (counts-as), and (ii) procedural information and enforcement artefacts are added to the institution.

The case study confirms the hypothesis, as we observed that conventional measures of physical access control, such as

physical security (walls, gates) and keys or ID passes for identification, authentication and authorization, count as “appropriate measures to prevent tampering with cargo units”. So the abstract concept of ‘tampering prevention’ was translated into the well known object of physical access control, and further refined into the enforcement artefact of security seals, for cargo units. Having to attach a seal onto a container is the addition of a concrete artefact, to enable enforcement of the open norm.

Hypothesis 5 (Normative beliefs): Norm emergence co-occurs with the creation of a normative belief in a community, i.e. a convention that the norm is generally going to be followed.

The case study shows that norms that are required for joining branch organizations or to get certified by standardization organization like ISO are more likely to be followed and thereby copied by other companies. This hypothesis is also confirmed. Mechanisms that motivate compliance (reduced administrative burden) may be developed for MAS to increase norm adherence.

4.1 Discussion

The research reported here has some limitations. In particular, the selection of research that we discussed is limited to normative multi agents systems. We are well aware that there is a lot of relevant other research about emergence (jurisprudence) and enforcement in the legal field, and in the social sciences.

Another limitation concerns the case study. Open norms make sense for safety and security in international trade, because the required security measures are so context dependent. In other areas, such as public health, food safety or fishing quotas, a strict enforcement regime may be more effective. We should not over-generalize findings from a single case.

A final limitation concerns the relevance to multi agent systems. Although ‘human’ solutions serve as a useful metaphor, artificial social systems may have their own laws.

5 CONCLUSIONS AND FUTURE WORK

In this paper we have present a case study of the adoption of AEO legislation in the Netherlands. The AEO guidelines are an example of open norms: norms of which the implementation is left open to interpretation. The case study demonstrates that in the case of open norms, (1) norm adoption, operationalization and implementation are guided in a top-down fashion, namely by EU guidelines and auditing standards, while (2) also emergence of norms, distribution and consensus reaching among companies, auditors and legal consultants, about the acceptability of implementations, are crucial. However, emergence always takes place within the constraints of the legal framework.

We believe that this dual nature is in fact common to many regulatory processes, including in particular the prescription, adoption and emergence of norms in artificial social systems and electronic institutions. On the one we expect that in such artificial societies, norms will be prescribed. On the other hand, not all behaviour can be described, and in practice, public decisions about what counts as acceptable behaviour will probably function as a boundary for future behaviour.

Which theories can account for this dual nature? We performed a survey of some of the literature in the multiagent systems field about norm emergence, norm specification and

norm adoption. First of all, most research about norm emergence studies the distribution of a given norm in a population or over a social network, not its creation. The research on cognitive agents and emergence does address this topic but is limited in their discussion of norm spreading. Second, research about the operationalization and instantiation of open norms only considers the reasoning process of individual agents. There is no account of the consensus building and copying behaviour, found in the case study. This shows that we need to make a combination: individual reasoning by subjects and regulators in order to generate possible interpretations of open norms, enhanced by communication processes to account for social acceptability of those possible interpretations.

REFERENCES

- [1] H. Aldewereld, F. Dignum, A.Garcia Camino, P. Noriega, J.A. Rodriguez-aguilar and C. Sierra (2006): Operationalisation of norms for usage in electronic institutions. In AAMAS (New York, USA, 2006), ACM Press, pp. 223–225.
- [2] B. Burgemeestre, J. Hulstijn, Y. Tan (2009): Agent Architectures for Compliance. In proceedings of the 10th International Workshop on Engineering Societies in the Agents World X, (ESAW 2009), Utrecht, The Netherlands
- [3] B. Burgemeestre, J. Hulstijn, Y. Tan (2009): Rule-based versus Principle-based Regulatory Compliance. In Frontiers in Artificial Intelligence and Applications; Vol. 205 Proceeding of the 2009 conference on Legal Knowledge and Information Systems (JURIX 2009), p. 37–46. Rotterdam, The Netherlands
- [4] R. Conte, G. Andrighetto, M. Campenni, M. Paolucci (2007): Emergent and Immigrant Effects in Complex Social Systems. In Proceedings of AAAI Symposium, Social and Organizational Aspects of Intelligence, Washington DC.
- [5] R. Conte, C. Castelfranchi (1999): From conventions to prescriptions. Towards an integrated view of norms. Artificial Intelligence and Law 7, p. 323–340
- [6] F. Dignum. Abstract norms and electronic institutions. In Proc. of the Int. Workshop on Regulated Agent-Based Social Systems: Theories and Applications (RASTA '02), Bologna, p. 93–104,
- [7] European Commission (2007) Authorised Economic Operators–Guidelines, TAXUD/2006/1450.
- [8] European Commission (2006) Authorised Economic Operators–The AEO Compact model, TAXUD/2006/1452.
- [9] Gribnau, H.: Soft law and taxation: The case of the Netherlands. *Legisprudence* 1(3) (2008)
- [10] Rukanova, B., Van Stijn, E., Henriksen, H., Baida, Z., Tan, Y.H. (to appear). Understanding the Influence of Multiple Levels of Government on the Development of Inter-organizational Systems. *European Journal of Information Systems*.
- [11] S. W. Sadiq, G. Governatori, and K. Namiri. Modeling control objectives for business process compliance. In G. Alonso, P. Dadam, and M. Rosemann, editors, *Business Process Management (BPM 2007)*, volume LNCS 4714, pages 149–164. Springer, 2007.
- [12] B.T.R. Savarimuthu and S. Cranefield, (2009) A categorization of simulation works on norms. In: Dagstuhl Seminar Proceedings 09121: Normative Multi-Agent Systems, 15–20 March, Leibniz, Germany, pp. 39–58.
- [13] Y. Shoham and M. Tennenholtz (1995). On social laws for artificial agent societies: online design. *Artificial Intelligence* 73(1–2):231 – 252.
- [14] J. Vazquez-Salceda, H. Aldewereld, and F. Dignum (2005): Norms in multiagent systems: From theory to practice. *International Journal of Computer Systems Science & Engineering*, 20(4):pages 225–236.

Non-taking up as a normative behaviour

Y. Saillard¹

Abstract. A part of the population could benefit from rights, in particular from social rights, but does not take advantage of them. This non-taking up behaviour is described and analysed using the normative multi-agent systems framework. The application of this approach gives an insight into the various components of this complex behaviour: the autonomy of the agents, their cognitive activity and the progressive construction of effective social rights.

1 INTRODUCTION

The notion of "Non-Taking Up" describes the behaviour of people who have some rights to social benefits or to public utilities but do not use them. Non-Taking Up reveals the distance between abstract and effective rights. This question is the object of numerous works which try to explain or to measure this phenomenon².

Some examples of Non-Taking Up in different fields can indicate the type of problems we aim at.

The Mexican antipoverty program Progresá (the Education, Health and Nutrition Program) began operating in 1997 in poor rural areas and has expanded to urban areas, [1]. It conditions monetary transfers to children's school attendance and regular health clinic visits of its beneficiaries families. From 1997 to 2003, three millions of households incorporated this Program. About 3 to 4 percent of the targeted families fail to take up the benefits. But, above all, 40 percent of children who are eligible for the education grants choose not to receive them.

In Greece and in Spain, non-take up of mean-tested retirement benefits is rather extensive [2]. According to different statistical estimates it seems that the non-taking up is higher the less automatic the claiming process is. For these benefits, non-taking up reaches 8 to 40 or even 60 percent in terms of the potential beneficiaries or of the potential expenditure. As a whole, 150,000 persons in Greece and 850,000 persons in Spain, of intended beneficiaries, fail to benefit from these programs.

'Standard Housing Benefit', a mean-tested social security payment, was intended to subsidise UK low income families, [3]. In 1981, about four millions of UK families received this benefit and two millions were entitled to claim this benefit and did not do so.

The examples are on mean tested benefits but the non-taking up behaviour is observed in many fields : minimum income guarantee, unemployment benefits, access to health care, recourse to justice, use of public and social services, [4,5,6,7,8].

The importance of non-taking up can be considered as revealing the inefficiency of social policy. Rates of non-taking up (non-taking up population/ eligible population) are often used as quantitative indicators. But these rates are not so easy to estimate and they oversimplify the individual and social processes at work [9]³.

Non-taking up cannot be viewed as the simple result of a binary choice of an isolated agent : taking/non-taking. The proper way to understanding non-taking up is to develop a comprehensive analysis of an agent who has some autonomy under social and institutional determinants and finally takes advantage or not of a social right or of a social benefit.

Many reasons can explain why potential rights are not used. Some public policies do not rightly appraise the populations affected and their needs. Beneficiaries may not receive enough information. Citizens, due to their past experience may decide of not using public utilities. Claiming to some social benefits or rights and appearing as beneficiaries can take a social meaning and produce a stigma which has a dissuasive effect.

As a first description of the processes leading to non-taking up, we propose the following diagram. The cognitive activity of agents and their social attributes in terms of representation of rights and of social membership lead to a progressive emergence of social right from an abstract goal to an effective right.

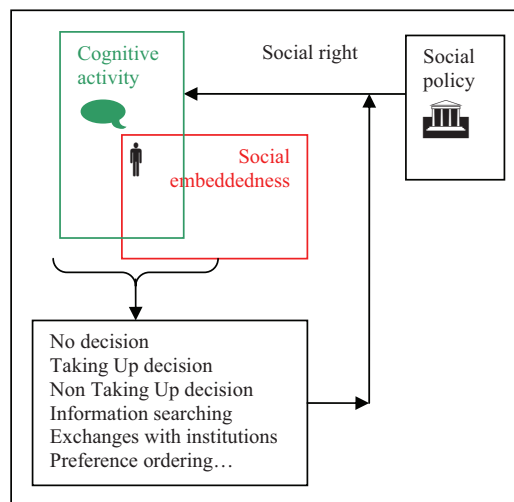


Figure 1. Individual behaviour and social rights : an overall process

¹ CNRS, LEPII, MSH-Alpes, University of Grenoble, BP 47 – 38040 Grenoble Cedex 9, France, yves.saillard@msh-alpes.fr.

² See, among others, [1], [2] and [3] for a discussion and a presentation of the notion of non-taking up. Extensive bibliographies are available at <http://odenore.msh-alpes.fr/biblio.html> and <http://www.exnota.org/bibliography.php>

³ The study of Regina T. Riphahn [9] is applied to the German social assistance program.

This paper extends works realized by the *Observatoire du Non Recours* (ODENORE) during the past five years⁴, and develops the analysis of the non-taking up process in terms of normative multi-agent systems.

The paper is organized as follows. First we characterize the non-taking up behaviours as appearing from empirical studies and observations applied to social policy. Next, in Section 3, we integrate this conceptual analysis in the normative multi-agent framework and we initiate an analysis of non-taking up in terms of normative positions. Section 4 presents our conclusions and provides directions for future work.

2 NON-TAKING UP BEHAVIOURS

Non-taking up arises from alternative causes that involve various types of behaviour, Table 1. This list is based on the research of EXNOTA and ODENORE aiming at a common definition of non-taking up, and on the synthesis they proposed of empirical studies [5,6,7].

Alternative causes of non-taking-up	Types of behaviour
<i>Individual Decisions</i>	
Demands for other forms of services (associative, private), conflict with social group membership, stigma,	Refusal of social assistance
Memory of past failure, negative learning	Individual learning on social policy
Negative utility, choice based on preference	Basic rationality
<i>Institutional Failures</i>	
Inefficient information, social policy failure	Unsatisfied demand
Inaccuracy of targeting population	Unsatisfied need

Table 1. Types of behaviour consistent with non-taking up

Even simplified, this list of types of non-taking up behaviours indicates that it can be consistent with very different situations, leading to an individual decision or to a lack of decision. At a first stage one may admit this distinction between "active" and "passive" non-taking up, with an active non-taking up thought as

a deliberate choice of exit from social services, and a passive non-taking up mainly due to institutions in large.

Meanwhile, our analysis of social rights as processes leads to reconsider this distinction. The distinction between "active" and "passive" non-taking up could mislead us in masking that social rights always imply actions from beneficiaries in most cases, to be effective. The very close tie between social rights over benefits and actions from the beneficiary is for us the main interest of referring to the normative positions approach to analyse the non-taking up behaviour (Section 3).

The non-taking up process, as briefly indicated in the Figure 1, must be more precisely defined. 'Rights to social benefits' become a reality depending on individual and social processes. Rights to social benefits do include rules which can be described as in expert systems, for instance :

```

If [(income < max level)
   and (min level < age < max level)
   and (residence in a given locality)...]
Then [(Social benefit = monthly amount "x")
      and (during "y" months)...]

```

But rights to social benefits need two other components to be defined: the information given to the population and the social meaning of the rule in terms of the social structure.

The information component of social rights is easy to understand. A potential beneficiary needs information to know what are his/her rights. This information must exist, be given, and must be clearly understandable. Above all, the realization of social rights implies lengthened exchanges of information between social institutions and beneficiaries: on the content of the social measures, on the situation of the beneficiary, with an alternation of the roles of supply and demand.

Social rights get their meaning with social embeddedness of the beneficiaries. Some social benefits may appear neutral for the social system if they do not create additional segmentation or discrimination. If they do discriminate populations, policies involve conflicts for the beneficiaries between the activation of these rights and their relationship to the social groups to which they belong. The conflict is all the more keen as their social group is deeply rooted in economic and social structures built in the long term.

The individual behaviour depends on the whole components of the social rights. So it cannot be a rational choice only based on monetary arguments. Individuals proceed to a complex cognitive activity. It includes a personal learning based on past experience, a calculus of utility, but it also takes an account of social implications of taking up social benefits. Many different situations can then be observed, from overuse of the available social rights to a refusal of any social benefit.

The cognitive activity of target populations aims at understanding the public supply and at building the terms of different answers to this supply. The 'activation' of social policies we can observe during the recent years, complexifies the public supply and enforces the possibility of non-taking up. It consists not only of adding more conditions for enjoying a social right (for instance, with mean-tested benefits), but also in fixing obligations of action (for instance, taking courses or receiving medical care). Alternatives opened to the individual and the requirements for public utilities affect this process. This whole cognitive activity puts the representations and the individual subjectivity into the foreground of the social policy efficiency. The autonomy of individuals and their ability to reason is

⁴ This observatory is first based on the results of exploratory research on the non-taking up of state services. It is a part of PACTE, laboratory at the Institute of Political Studies of Grenoble. Odenore is localised at the Maison des Sciences de l'Homme Alpes. As head of the European thematic network EXNOTA (Exit from and non-take-up of public services) created in 2003 with the support of the 5th Research and Technological Development Framework Programme of the European Union, ODENORE leads a multidisciplinary research team working on non-taking up, integrating politics, sociology and law. Philippe Warin (Director of Research CNRS, PACTE/CERAT) is the Scientific Director of the Team.

progressively built from their past experience and from mutual interactions.

'Cognitive activity' does not mean a hyper rationality. Individuals can base their representation and their behaviour on routine, imitation, without clear knowledge, due to public supply or to their own difficulties.

The whole process develops with a time dimension. Past experience and individual learning are important elements of the process. But the construction of the demand is also progressive and needs exchanges between individuals and institutions to declare for a demand, giving the necessary information, fitting the benefit to the individual case...

In addition to the social policy institutions and to the eligible population, some agents can contribute to the process of realization of social rights. A third party such as mutual aid associations can positively activate the process. Here, we do not treat of this point.

The observations describe a large diversity of the forms of non-taking up and rather than simply opposing taking up and non-taking up, we have to specify behaviour according to multiple criteria: duration, intensity, adding or not to other refusals, compensated or not by alternative supplies.

Here, our attempt to use the normative multi-agent systems for a better understanding of non-taking up does not aim at proposing a complete analysis of the complexity of the questions we have pointed. Our main aim is to demonstrate that this approach is relevant, the analysis having to be specified for definite cases of social rights or benefits.

3 NON-TAKING UP AND NORMATIVE MULTI-AGENT SYSTEMS

Autonomy of agents with regard to social policy is the key feature of non-taking up. It is also the central question for the normative multi-agent systems. We first consider how the general framework produced for multi-agents systems and norms, is adapted for analysing non-taking up. Then we apply more specifically the concept of normative position.

3.1 Non-taking up and autonomy of normative agents

The analysis of social rights in terms of normative multi-agent systems implies a slight conceptual shifting. Norms are generally treated as constraints on the human autonomy, necessary to overcome problems of coordination. But social rights are rather a mean of extending the capacity of agents and of increasing their capacities. Then, non-taking up can be seen as a refusal or the non-achievement of this opportunity.

For standard economists, a non-taking up behaviour could be only explained by costs associated with the use of social rights. It could be the consequence of the preference ordering of the eligible agent, or, as in the new institutional economics, of public failure in the supply of services. But in any case, non-taking up would be based on monetary terms⁵.

It seems that the treatment of the norms in agent-based systems can make way to a more satisfactory analysis of the non-taking up question. Here, we will use the framework proposed by Fabiola López y López, Michael Luck, and Mark d'Inverno, [11], coming from SMART.

The main object of this framework is not to describe how norms are created, but to clarify for different types of norms what are the processes that explain the normative behaviour of agents. Rightly, the authors emphasize that this process cannot be reduced to a simple accounting of gains and losses by each agent.

Within the SMART framework, norms are defined in a general way. They have three main components: prescriptiveness (how an agent has to behave), sociality (an agent among other agents), social pressure (restraints on the individual behaviours), [11]. Their main social function is to specify patterns of behaviour through normative goals directed to addressee agents.

Taking up a benefit can be identified with a behaviour of instancing a norm, which means for an agent that the norm is adopted and internalised by him. It implies for this agent changing his/her pattern of goals. As for norm acceptance [12], this choice is based on autonomous decisions which induce normative goals of agents. The distinction between external and internal sides of the norms [13] is relevant, with the visible display of the social policy and its internal representation and belief by the agents.

Normative agents are able to act on norms and to reason about them. The addressee agents must be members of the system but do not know or adopt every norm, like eligible people for social rights or benefits. They are free to adopt some norms, then to comply with some of them, and to reject others. For our own concern, of the different concepts of the SMART agent framework, one is of particular interest, sets of capabilities which indicate what agents are able to perform.

The multiplicity of norms (with particular cases of interlocking) may involve internal conflict for agents for deciding which norms they have to comply with, [14]. Then, autonomous reasoning permits agents to choose between contradictory objectives. What is important here is that constraints do not directly affect actions but the patterns of goals which inspire individual actions, [15]. Within the SMART framework, the agents choose how to act after weighing up their preferences between alternative goals, punishments for not complying with a norm or rewards for complying with it.

The acceptance of a social benefit may affect how the own goals of the beneficiary can be achieved (owing to, for example, an improvement of resources). Thus, it also means that without social policy, these goals would be achieved differently. Benefit acceptance strengthens authority and legitimacy of social institutions. Moreover, conditional social policies can force agents to follow goals different from their own goals. Analogous with institutionalised power [16], 'active' social policy aims at changing goals of beneficiaries. Then, with the autonomous decisions of agents, the way for non-taking up is widened.

Regarding the social meaning of benefits and the eventuality of stigma, the question is to combine the direct returns on individual preferences and the effects of social policy on the relationship between the beneficiary and the groups to which she/he belongs. For example, G.Akerlof analysed social programs such as educational attainment and childbearing and how group interactions and social externalities have an influence

⁵ See the study of Janet Currie [10] for a review of literature regarding the take up of social programs in the U.S and U.K., with an economic point of view.

on individual decisions [17]. It amounts to specify for social policy the meaning of social membership, [18].

In the SMART conceptual framework, instancing a norm is the sign of the transition from an abstract norm to its effectiveness. Instancing is an individual process in that each agent works on the norm and appropriates it. This conceptualisation of the implementation of norms is in accordance with what is observed for social policies and with the notion of coproduction of public services by the citizens⁶.

The splitting up of this whole reasoning process by the SMART framework is well suited. We can also consider that there is a first process (like the "norm adoption process") through which the agents recognize (or do not recognize) that they are concerned about some target benefit. It implies not only to recognize the legitimacy of social policy, but it also implies that the agent admits to be eligible (the equivalent here of being an "addressee" in the norm adoption process) according to the criteria used for the delivery of the benefit. Then the deliberation process consists of estimating how the pattern of goals is modified due to positive opportunities but also due to restraints associated with the benefit. The SMART framework includes norms of enforcement, reward and penalties, which in fact are parts of the appliance of social policy. This deliberation is based on the agent's motivations. Finally, the beneficiaries have to actualize their goals (the "norm compliance process"): the goals they pursue and the priority between their goals.

3.2 Normative positions

The concept of "Normative positions" comes from a whole series of works whose common objective is to proceed to a systematic analysis of the behaviours we can expect from the agents concerned with norms. The analysis of normative positions refers to the right analysis and to the founding theory of Höhfeld. But if this approach originates in the analysis of right, it aspires to apply to the whole interactions between agents [20]⁷.

The normative positions approach characterizes elementary types of behaviour which are specified in terms of obligations, permissions, rights and duties, commitments, authorizations, responsibilities. But the main contribution for our own concern is the analysis of the internal consistency of behaviour of a given agent and the consistency with behaviour of other agents involved in a given normative situation.

We will use the presentation of the Kanger-Lindahl theory and the notation proposed by Marek Sergot [20,21] for a preliminary presentation of the social benefit relation. We follow the first steps of an analysis like that used by A. Jones and M. Sergot for security requirements in the access to computerised database of patients [22], and by M. Sergot for loan of books, the problems coming from a fence on the boundary between adjoining properties, and for car park, [20].

There are two agents: "e" (any agent eligible for a given benefit), and "g" (the agent who governs the social policy), and a fact "B" (the delivery of the benefit). With the normative one-agent act positions defined by Lindahl, the social benefit relation is represented as the following set of positions:

$$[\pm P[\pm E_e \pm B]], [\pm P[\pm E_g \pm B]]$$

With the notations used by Marek Sergot, this scheme means that [It is permitted, or not [that "e", respectively "g", sees to it or not that B be true or that B not be true]]. We also use the abbreviation proposed by Lindahl and Marek Sergot of the passivity of agent "e":

$$\text{Pass}_a B = \neg E_a B \wedge \neg E_a \neg B$$

The interpretation of the seven Lindahl's normative one-agent ("a" is "e" or "g") act positions, is the following

(T₁) $P E_a B \wedge P E_a \neg B \wedge P \text{Pass}_a B$: "a" is permitted to see to it that the benefit is delivered, to see to it that the benefit is not delivered, and to be passive;

(T₂) $P E_a B \wedge O \neg E_a \neg B \wedge P \text{Pass}_a B$: "a" is permitted to see to it that the benefit is delivered and to be passive, and has the obligation to not see to it that the benefit is not delivered;

(T₃) $P E_a B \wedge P E_a \neg B \wedge \neg P \text{Pass}_a B$: "a" is permitted to see to it that B is delivered and that B is not delivered, and is not permitted to be passive;

(T₄) $O \neg E_a B \wedge P E_a \neg B \wedge P \text{Pass}_a B$: "a" is obliged not to see to it that B is delivered, and is permitted to see to it that the benefit is not delivered and is permitted to be passive;

(T₅) $O E_a B$: "a" is obliged to see to it that the benefit is delivered;

(T₆) $O \text{Pass}_a B$: "a" is obliged to be passive;

(T₇) $O E_a \neg B$: "a" is obliged to see to it that B is not delivered.

How do these preliminary expressions let a place to non-taking up? We will examine the meaning of these expressions first for the eligible agent "e", then for the social policy organizer "g".

The eligible agent "e", is obliged, with the normative act position T₄ not to see to it that B is delivered. T₇ would be a nonsense: "e" would be obliged to see to it that B is not delivered. The obligation to be passive, T₆, seems also unrealistic. There is no more place for non-taking up if the normative eligible agent act position is the obligation to see to it that B is delivered, as with T₅.

The discussion about non-taking up as it is approached with the Lindahl's list, concentrates, from the eligible agent point of view, round three expressions: T₁, T₂, and T₃. T₁ describes a position of complete freedom for the eligible agent, leading to the different forms of non-taking up. T₂ puts the autonomy of the eligible agent under constraint: there is no alternative to social policy for the agent. The eligible agent loses the capacity to act for the non-delivery of the benefit for a capacity of no acting for the delivery of the benefit or to be passive. T₃ is a normative position which does not permit the passivity, but the eligible agent can either see to it that B is delivered or see to it that B is not delivered.

From the point of view of the other generic agent "g", the organizer of the social policy, the expressions rank in a different manner. For this agent, there is an obligation to deliver the benefit. It excludes expressions which would permit "g" to act against the delivery of the benefit, like with T₁, T₃ or T₄. We cannot admit that the organizer of social policy would have the obligation to be passive (normative act position T₆) or to see to it that the benefit is not delivered (T₇).

⁶ For instance, the contribution of G.P. Whitaker, [19].

⁷ Precisely: "all forms of regulated and organised agent interaction, legal or non-legal, formal or informal", p. 581.

So, for the social policy maker "g", the discussion about non-taking up concentrates on two expressions: T_2 and T_5 . T_5 would correspond to the simple obligation for "g" to see to it that the benefit is delivered. T_2 implies that "g" cannot do anything which could prevent the delivery of the benefit, but "g" might be passive, and the obligation to see to it the delivery of B disappears.

This preliminary discussion leads to a typology of non-taking up based on the normative positions of the eligible agent and of the social policy maker:

Case 1 $P E_e B \wedge P E_e \neg B \wedge P Pass_g B$ $O E_g B$
Case 2 $P E_e B \wedge P E_e \neg B \wedge P Pass_g B$ $P E_g B \wedge O \neg E_g \neg B \wedge P Pass_g B$
Case 3 $P E_e B \wedge O \neg E_e \neg B \wedge P Pass_e B$ $O E_g B$
Case 4 $P E_e B \wedge O \neg E_e \neg B \wedge P Pass_e B$ $P E_g B \wedge O \neg E_g \neg B \wedge P Pass_g B$
Case 5 $P E_e B \wedge P E_e \neg B \wedge \neg P Pass_e B$ $O E_g B$
Case 6 $P E_e B \wedge P E_e \neg B \wedge \neg P Pass_e B$ $P E_g B \wedge O \neg E_g \neg B \wedge P Pass_g B$

Table 1. Non-taking up and normative positions

We can organize the different cases in relation with possible non-taking up (Diagram 2). The point TU indicates a conjunction of positions with no place for non-taking up. The basis of the diagram corresponds to a normative act position of the policy maker agent "g" who has the obligation to act for the delivery of the benefit. The upper face of the diagram corresponds to the removal of this obligation and we consider it is leading to inefficiency from the governmental institutions. The vertex of the lower and of the upper faces correspond to the conjunction of different normative act positions of the eligible agent "e", with respectively obligation and no obligation for the policy maker to see to it that B is delivered. Taking TU as a point of departure, non-taking up behaviour might be due to preference ordering and to a low involvement in social policy of the eligible agent "e" (Case 3). Non-taking up might also come from alternatives to B and from an exit behaviour of "e", (Case 5). Case 1 holds the different forms of non-taking up due to preference ordering, alternatives-exit and low involvement.

Cases 4, 6 and 2, add respectively to Cases 3, 5 and 1, the inefficiency coming from the offer of benefit.

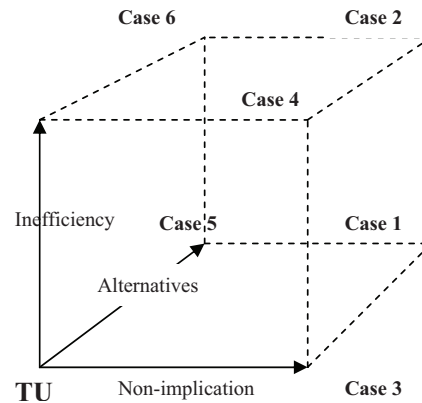


Diagram 2. Forms of non-taking up and normative positions

The specification of normative positions, even based on a preliminary approach, involves a clarification of the content of measures of social policy. Who is concerned? What are the constraints associated with the benefit? What are the actions that an eligible agent needs to do in order to realize a potential right. But, above all, simple assumptions lead to distinguish cases of non-taking up near those drawn from empirical studies, as it appears on the Table 1. It puts forward some decisive processes such as: individual preferences, adhesion to policies, alternative to benefits, insufficiency or inefficiency of the offer.

The Diagram 2 shows the beginnings of collective dynamics initiated by a measure of social policy, the implications of this measure on the actions of concerned agents according to its terms, and the regulated effects on the states of the system.

The agent's choice is determined from existing states, his preference ordering and the constraints on his actions. The specification of normative positions contributes to define which actions lead to permissible states. Then we have to return to the architecture of multi-agent system and to specify how it integrates the normative system, [23]⁸.

4 CONCLUSIONS & FUTURE WORKS

The normative multi-agents systems and the theory of normative positions constitute a fruitful framework in the analyse of autonomous behaviour such as non-taking up of social rights or benefits.

The general architecture of normative multi-agents is equipped through the transformation of goal patterns and preference ordering associated to normative situations, in order to contribute to the understanding of refusal, compromise, and acceptance. Even if it implies an evolution of the conception of

⁸ DALMAS is a proposition for such an architecture for norm-regulation of multi-agent system, and more precisely for the interplay between preferences and norms.

norms as constraints. In the broad sense of the concept, they can also extend the individual objectives and capacities.

The practicable application of this approach implies to go deeply into several points.

First we have to analyse systematically how the numerous empirical studies describe the non-taking up/taking up behaviour. Surveys and synthesis exist, as those cited at the beginning of the paper. But we have to stylize these results with the general architecture of normative multi-agent systems in the background.

Non-taking up involves individual positions and behaviours which are in relation with social determinants. Institutions, representations, stigma, conflict of norms,... refer to agents in a society and to members of social groups. Goal patterns are linked to social roles. Then, the social organization of the agents must also be explicitly introduced in the architecture of the system.

Social policy is much more complex than only giving or not a benefit or a right to do or to get something. The measures are composites or various elements: rules, advice, service, control. Several types of agents contribute to the realization of these measures, for instance, with regard to health benefits : insurance institutions, physicians, intermediary associations and non governmental organizations. Thus, in addition to the inherent complexity of the method of analysis, we need an efficient strategy for a progressive comprehension of the area.

Moreover, the application of normative behaviours to non-taking up and to the research of a better efficiency of social policy would need some automated tool. It is closely akin to the objective fixed by Marek Sergot to the system "Norman-G" used to apply the theory of normative positions and to analyse practical examples, [20]. It is also the objective of works based on the DALMAS architecture, [23]. Such a tool is necessary for an extension to more than two agents and for a more realistic description of social policy institutions. It is also necessary for a systematic analysis of the implications of the rules and of their terms.

REFERENCES

- [1] S.W. Parker and G.M. Teruel. Randomization and Social Program Evaluation: The Case of Progres. *Annals of the American Academy of Political and Social Science*, Place Randomized Trials: Experimental Tests of Public Policy, 599:199-219 (2005).
- [2] M. Matsaganis, H. Levy, and M. Flevotomou. Non take up of means-tested benefits for the elderly in Greece and Spain. *7th Espanet Conference 2009*, session stream nr. 10 - Poverty risks and social policy programmes in the 21st century the future of the welfare state: paths of social policy innovation between constraints and opportunities. Urbino 17-19 September, 17p. (2009).
- [3] R. Blundell, V. Fry, and I. Walker. Modelling the Take-up of Means Tested Benefits : the Case of Housing Benefits in the United Kingdom. *The Economic Journal*, 98(389):58-74 (1988).
- [4] V. Hernanz, F. Malherbet, and M. Pellizzari. Take-up of Welfare Benefits in OECD Countries : a review of the Evidence. *OECD social, employment and migration working papers*, 17, 47 p. (2004).
- [5] Exnota. *Exit from and non-take up of public services. A comparative analysis: France, Greece, Spain, Germany, Netherlands, Hungary. (Emergence of "non-take up")*. Final Report Exnota Consortium, http://www.exnota.org/pdf/emergence/final_report_emergence.pdf. 30p. (2005).
- [6] Exnota. *Exit from and non-take up of public services: A comparative analysis: France, Greece, Spain, Germany, Netherlands, Hungary. (Measurement of "non-take up")*. Final Report Exnota Consortium, http://www.exnota.org/pdf/measurement/final_report_measurement.pdf. 57p. (2005).
- [7] ODENORE. *Non-taking up: Methodological Preconditions and Works*. C. Chauveaud, I. Costerg, P. Warin, <http://odenore.msh-alpes.fr/travaux.html>, last modified on 15th october 2009 (2009).
- [8] W. Van Oorschot and J.H. Win. Non-take-up of social security benefits in Europe. *Journal of european social policy*, 1(1):15-30 (1991).
- [9] R. T. Riphahn. Rational poverty or poor rationality ? The take up of social assistance benefits. *Review of Income and Wealth*, 47(3):379-398 (2001).
- [10] J. Currie. The take-up of Social Benefits. *IZA Discussion Paper Series*, 1103, 51p. (2004).
- [11] F. López y López, L. Luck and M. D'Inverno. A normative framework for agent-based systems. *Computational Mathematical Organisation Theory*, 12:227-250 (2006).
- [12] R. Conte, C. Castelfranchi, and F. Dignum. Autonomous norm-acceptance. In *Intelligent Agents V (ATAL'98)*. J. Müller, M. Singh, and A. Rao (Eds), LNAI 1555, Springer-Verlag, pp. 319-333 (1999).
- [13] R. Conte and C. Castelfranchi. Norms as mental objects. From normative beliefs to normative goals. In C. Castelfranchi and J. P. Müller (Eds), *From Reaction to Cognition (MAAMAW'93)*, LNAI 957, Springer-Verlag, pp. 186-196 (1995).
- [14] F. López y López and L. Luck. Modelling Norms for Autonomous Agents. In: *Proceedings of the Fourth Mexican International Conference on Computer Science (ENC'03)*. E. Chávez, J. Favela, M. Mejía, A. Oliart (Eds). IEEE Computer Society, pp. 238-245 (2003).
- [15] F. López y López and L. Luck. A model of normative multi-agent systems and dynamic relationships. In: *Regulated agent based social systems*. G. Lindemann, D. Moldt, M. Paolucci (Eds). LNAI 2934, Springer Verlag, pp. 259-280 (2004).
- [16] A. Jones and M. Sergot. A formal characterisation of institutionalised power. *Logic Journal of the IGPL*, 4(3):429-445 (1996).
- [17] G.A. Akerlof. Social distance and Social Decisions. *Econometrica*, 65(5):1005-1027 (1997).
- [18] R. Conte and C. Castelfranchi. *Cognitive and Social Action*. UCL Press, (1995).
- [19] G.P. Whitaker. Coproduction: Citizen participation in service delivery. *Public Administration Review*. 40(3):240-246 (1980).
- [20] M. Sergot. A Computational Theory of Normative Positions. *ACM Transactions on Computational Logic*, 2(4):581-622 (2001).
- [21] M. Sergot. Normative Positions. In: *Norms, Logics and Information Systems*. P. McNamara, H. Prakken (Eds). IOS Press, Amsterdam, pp. 289-308 (1998).
- [22] A. J. I. Jones and M. Sergot. Formal Specification of Security Requirements using the Theory of Normative Positions. In: *Computer Security - ESORICS*. Y. Deswarte, G. Eizenberg, J.-J. Quisquater (Eds). Lecture Notes in Computer Science, 648, Springer-Verlag, Berlin Heidelberg, pp.103-121 (1992).
- [23] J. Odelstad and M. Boman. Algebra for agent norm-regulation. *Annals of Mathematics and Artificial Intelligence*. 42:141-166 (2004).

An Analysis of Norm Emergence in Axelrod's Model

Samhar Mahmoud¹ and Nathan Griffiths² and Jeroen Keppens¹ and Michael Luck¹

Abstract. Norms are a valuable mechanism for establishing coherent cooperative behaviour in decentralised systems in which no central authority exists. One of the most influential formulations of norm emergence was proposed by Axelrod [1], who defined a model of norms and metanorms that enables norm establishment in populations of self-interested individuals. This paper provides an empirical analysis of aspects of Axelrod's approach, by exploring some of the key assumptions made in previous evaluations of the model. First, we explore the dynamics of norm emergence and the occurrence of norm collapse when applying the model over extended durations. Second, we investigate in detail the reasons for norm collapse in extended applications of the model and show that both the level of mutation in the population and the precise nature of the reproduction mechanism are significant. Our findings identify characteristics that significantly influence norm establishment using Axelrod's formulation, but are likely to be of importance for norm establishment more generally.

1 Introduction

As has been suggested by many (e.g., [2, 3, 4, 8, 12]) norms provide a valuable mechanism for regulating or constraining human societies. Perhaps the most obvious and clear manifestation of norms is when they arise through the explicit introduction of laws that are established by legislatures, for example, or through rules or bye-laws of smaller groups such as member clubs. However, norms are also valuable when there is no central authority, and they emerge as a result of individual behaviour, in order to establish some coherence or stability in a group. It is this latter aspect that has been the focus of several researchers (e.g., [5, 6, 10, 13, 11]) perhaps most notably Axelrod, whose seminal paper in 1986 offered a model of norms and metanorms [1] that has since been further investigated [7, 9].

Axelrod's model is a game in which different agents decide whether to defect or cooperate (comply). Agents may also observe others and have the ability to punish those who defect. An agent's behaviour is assessed by means of a careful system of scoring that simulates the potential rewards and penalties associated with norm violation and enforcement. The agent population is evolved through a number of iterations, with a mechanism whereby successful behaviour (as measured by the scoring system) tends to be replicated and unsuccessful behaviour tends to be discarded. In each iteration, each replicated behaviour is subjected to a small chance of mutation, reflecting the feature that an agent may occasionally change its strategy, irrespective of past habits. The strategy of each agent in determining whether to defect and whether to punish others is determined by two different attributes, *boldness* (encouraging agents to

defect) and *vengefulness* (encouraging them to punish others), which are distinct for each agent. The idea is that a system eventually resulting in all agents having high vengefulness and low boldness corresponds to norm emergence, since they will punish defection but they will not themselves defect. Key to Axelrod's model is the notion of *metanorms*, secondary norms that help to enforce compliance with primary norms by punishing agents that do not themselves punish a defector. By using metanorms, Axelrod was able to establish norms in his experiments.

However, as was more recently shown by Galan and Izquierdo [7], Axelrod's results are dependent on both certain assumptions and some very specific and arbitrary conditions. In this paper, we elaborate on the work of Galan and Izquierdo, showing that their results, too, rely on some assumptions and conditions, and we provide a further analysis of Axelrod's model, drawing out some important considerations for the establishment of norms more generally.

The paper begins with a description of Axelrod's *Norms Game*, followed by a more detailed analysis of the results than provided elsewhere. Then, in Section 4, the duration of the game, a critical part of Galan and Izquierdo's analysis, is reviewed, with different results, leading to a new consideration of the circumstances for norm collapse (when norms are not established). In Section 4, the impact of the reproduction policy is analysed, and in Section 5, a consideration of the impact of mutation is provided. The paper concludes with a discussion and conclusions on the significance of these results.

2 Axelrod's Model

2.1 The Norms Game

Axelrod's *norms game* adopts an evolutionary approach in which successful strategies multiply over generations, potentially leading to convergence on norms. Each individual, or agent, can choose to *defect* by violating a norm, and such behaviour has a particular known chance of being observed, or *seen* (S). An agent i has two decisions, or strategy dimensions, as follows. First, it must decide whether to defect, determined by its *boldness* (B_i); and second, if it sees another agent defect (determined by S) it must decide whether to punish this defecting agent, determined by its *vengefulness* (V_i), which is the probability of doing so. If $S < B_i$, then i defects, receiving a *temptation payoff*, $T = 3$, while *hurting* all the others with payoff $H = -1$. If a defector is *punished* (P), the payoff is $P = -9$, while the punishing agent pays an *enforcement cost* $E = -2$. The initial values of B_i and V_i are chosen at random from a uniform distribution of a range of eight values between $\frac{0}{7}$ and $\frac{7}{7}$.

Axelrod's simulation has a population of 20 agents, with each agent having four opportunities to defect, and the chance of being seen for each drawn from a uniform distribution between 0 and 1. After playing a full round (all four opportunities), scores for each agent are calculated in order to produce a new generation, as follows.

¹ Department of Computer Science, King's College London, London WC2R 2LS, United Kingdom. email: samhar.mahmoud@kcl.ac.uk

² Department of Computer Science, University of Warwick, Coventry CV4 7AL, United Kingdom.

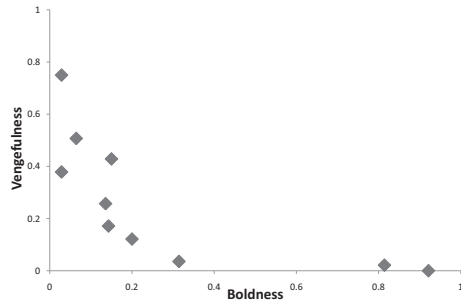


Figure 1: Norms game overall results

Agents that score better or equal to the average population score plus one standard deviation are reproduced twice in the new generation. Agents that score one standard deviation under the average population score are not reproduced at all, and all others are reproduced once. Although this may produce a new generation with a different number of agents, Axelrod maintains the number of agents at 20 over subsequent generations, but does not specify how. Finally, a mutation operator is used to enable new strategies to arise. Since B_i and V_i (which determine agent behaviour) take 8 possible values, they need three bits to be represented, to which mutation is applied (by flipping a bit) whenever an agent is reproduced, with a 1% chance.

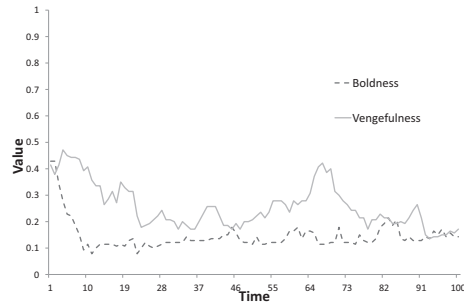
Axelrod's experiment was run five times, each with 100 generations. Two runs resulted in high average boldness and almost zero average vengefulness so there was no norm emergence at all, two other runs gave low boldness and vengefulness, but only the final run had a high level of vengefulness and very low boldness, indicating a partial establishment of a norm against defection. In response, Axelrod considered an additional mechanism to support norms in his *metanorm* model.

2.2 The Metanorms Game

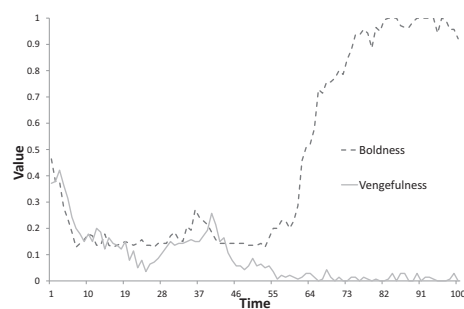
The key idea underlying Axelrod's metanorm mechanism is that some further encouragement for enforcing a norm is needed. This is accomplished by introducing a *metanorm* for punishment of those who observe a defection but do not punish the defectors. In this new metanorms game, if an agent sees a defection but does not punish it, this is considered as defection itself, and others in turn may observe this defection (with probability S) and apply a punishment to the non-enforcing agent. As before, the decision to punish is based on vengefulness, and brings the defector a punishment cost of $P' = -9$ and the punisher an enforcement cost of $E' = -2$. Applying this new metanorm game to the same simulation as before gives runs with high vengefulness and low boldness, which is exactly the kind of behaviour needed to support establishment of a norm against defection.

3 Analysis of Axelrod's Model

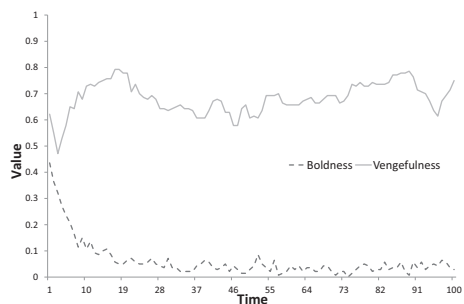
In seeking to replicate Axelrod's results [1], it becomes clear that some assumptions need to be made, about which Axelrod says nothing. First, the model does not specify how the constant population level is maintained after reproduction, when there are three possible scenarios. (i) The new population is smaller than the original. In our re-implementation, additional agents are randomly selected from the resulting population and replicated. (ii) The new population is equal in size to the original, in which case nothing new is needed. (iii) The



(a) Norms game: low boldness and low vengefulness



(b) Norms game: high boldness and low vengefulness



(c) Norms game: low boldness and high vengefulness

Figure 2: Norms game: analysis of runs

new population is larger than the original. In this case, our position is to determine the required number of agents at random from the relevant set for reproduction. Second, we assume the score of each agent is set to 0 at the beginning of each generation.

We repeated Axelrod's experiments, running the norms game 10 times, with results as shown in Figure 1, where the diamonds represent the value of the mean average boldness and vengefulness of the final generation population. This is similar to Axelrod, with one run having high vengefulness and low boldness, two runs with exactly the opposite (high boldness and low vengefulness), and all other runs with low values for both boldness and vengefulness.

In order to establish how these results arise, changes to boldness and vengefulness for each individual were monitored; Figure 2 provides some sample graphs illustrating this, showing the average boldness and vengefulness as they vary over generations (and hence time). In particular, Figure 2(a) shows one run ending in the

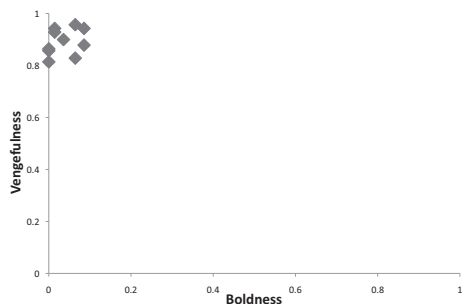


Figure 3: Metanorms game overall results

most common result, low boldness and low vengefulness. The run starts with average boldness and vengefulness of about 0.5 (as initial values for B_i and V_i are taken from a uniform distribution over $\{\frac{0}{7}, \dots, \frac{7}{7}\}$). In the early stages, boldness decreases slightly, indicating that individuals with higher boldness are eliminated. This is because high boldness causes an agent to defect, yet defecting with average vengefulness can be costly, as the agent is likely to be punished, leading to a low score. Subsequently boldness stabilises at a low level. Finally, the values stabilise at particular low values for both boldness and vengefulness until the end of the run.

In the cases that result in high boldness and low vengefulness (an example run is shown in Figure 2(b)), the run starts as before, with both values reducing. However, around the 60th generation, the value of boldness increases sharply until it reaches 1, where it remains. This can be explained by a dramatic change to one individual's boldness, due to mutation, in an agent population with particularly low values of vengefulness. In turn, this facilitates the individual's survival, dominating the others and allowing it to propagate its high boldness across the population. Here, a high score is attained by defecting without punishment (due to low vengefulness), which also hurts others and lowers their scores.

In the final case, as shown in Figure 2(c), the run ends with high vengefulness and very low boldness: when the high boldness phase ends, only individuals with high vengefulness remain, so there are no individuals with low boldness and low vengefulness, and those with high vengefulness and low boldness survive and dominate.

By introducing metanorms, Axelrod aimed to address the problems identified above. Our own simulation (see Figure 3) provides similar results to Axelrod but again, a deeper analysis is required. As shown in Figure 4, in this new game, the population starts eliminating low boldness individuals as before, but now also eliminates low vengefulness individuals. The latter trend results from the metanorm game in which failure to penalise a defector may also be penalised. This results in a population with high vengefulness and low boldness, which survives until the end of the run.

4 Game Duration

In analysing the model, we repeated these experiments over a *long duration*, 1,000,000 generations (and 10 runs), as opposed to 100 generations, to provide a stronger analysis. In our norm game simulation, the game starts with boldness decreasing, and then vengefulness decreasing until they both settle at a low level, which is consistent with Axelrod's results. However, an agent with high boldness can be introduced to such a population through mutation, and would domi-

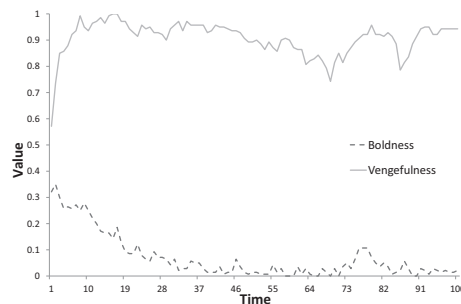


Figure 4: Metanorms game analysis

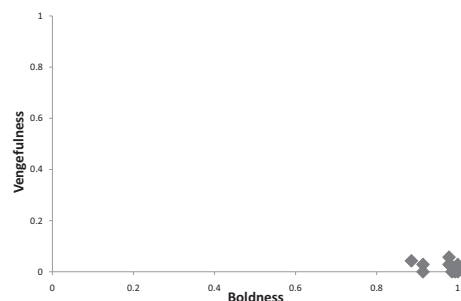


Figure 5: Norms game for 1,000,000 generations

nate others since it is not punished due to low levels of vengefulness. Clearly, running the experiment for a longer period increases the opportunities for this to occur and, as shown in Figure 5, this always leads to norm collapse.

In undertaking their own analysis of Axelrod's metanorm model, Galan and Izquierdo [7] increased the number of generations in a run and found different results. By including 1,000,000 generations, in 1,000 runs, nearly 70% ended in *norm collapse*, as opposed to Axelrod's *norm establishment*. According to Galan, this is because vengefulness is costly in a population in which violation is rare. Thus, agents with low vengefulness are favoured over agents with high vengefulness, leading to a significant decrease in vengefulness, encouraging defection, and in turn causing boldness to increase. Galan's results suggest that metanorms are not as useful as it might seem from Axelrod's results.

In analysing these cases to determine the reasons for this result, it's clear that the runs begin in the same way as previously observed, by eliminating individuals with high boldness and low vengefulness, stabilising on those with high vengefulness and low boldness. Then, however, mutation causes vengefulness to reduce. If an agent *A* with high vengefulness and low boldness changes through mutation to give lower vengefulness, while boldness for all remains low, there is no defection and the mutated agent survives. In addition, if boldness then mutates to become just a little higher for a different agent *B*, with average vengefulness remaining high, *B* will still rarely defect because of relatively low boldness. If it *does* defect, and is seen by others, it receives a low score, unless it is not punished, in which case the non-punishing agents may themselves be punished because of the high vengefulness in the general population. Here, agent *A* may

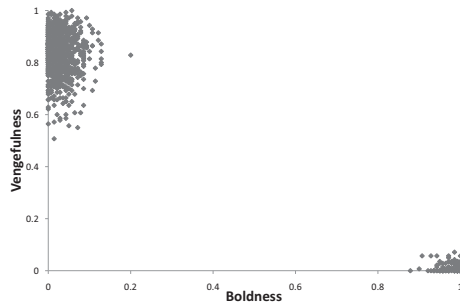


Figure 6: Metanorms game for 1,000,000 generations

not punish B because of the low probability of being seen (which must be below the low boldness level to have caused a defection) or because it has mutated to have lower vengefulness. In the former case, A will not be punished for non-punishment, but in the latter case, A might be if it is seen by others. However, we know that the probability of being seen is low because agent B has defected (and $S < B$ for defection to take place). In this case, B is eliminated, while A remains, because the likelihood of B 's defection being seen by just one agent is relatively high, but the likelihood of agent A 's non-punishment being seen requires first B 's defection being seen by A , and then A 's non-punishment being seen by others, the combination of these being extremely unlikely.

If the values of vengefulness continue to decrease in this way, the population can arrive at a situation with very low average boldness and vengefulness. At this point, a single mutation to boldness could then cause the mutant to dominate the others due to the general lack of vengefulness in the population. The key question here is why, in cases of high boldness and low vengefulness, mutation of vengefulness from a very low value to a significantly higher value does not cause boldness to decrease. Here, such a mutant should punish all others for defecting *and* for not punishing defectors. However, these punishments also incur significant enforcement costs, all of which are borne by the punishing agent, potentially exceeding the penalty meted out to the defectors and those agents who fail to punish others.

This analysis suggests that both mutation and sanctioning play a major role in collapsing norms. However, there is an additional factor that gives rise to these results, a particular characteristic of the underlying model which, in certain circumstances, and with very subtle change, can give a very different outcome. We consider this next.

5 Reproduction and Norm Collapse

As specified earlier, a run of the metanorm game settles at very low boldness and very high vengefulness at a certain point. For this to change to the opposite situation of low vengefulness and high boldness, a sequence of modifications that lower vengefulness must occur, and another sequence of modifications that increase boldness must also occur. At this point, the population stabilises at the same level of vengefulness and boldness until the end of the run. Now, when all individuals have boldness around 0, defection rarely happens, and their scores (which change only when agents are hurt, punished or enforce) are 0. As a result, the average score and standard deviation are also 0, so that all agents have a score equal to the average score plus one standard deviation. This means that all agents are replicated and then the new generation is selected at random from

this pool of agents, as described in Section 3.1

It is this key point that is critical to these striking results. According to Axelrod's rules, agents in this situation should be replicated *twice* when forming the new generation, because all agents have a score equal to the average plus one standard deviation. However, duplicating the individuals in this case does not seem sensible since it does not fulfil the original purpose suggested by Axelrod, of giving individuals with better scores a greater chance of survival. If, in contrast, as we have done in our own simulation for this special case alone, we only replicate once, then the results of the metanorm game with many more generations are similar to those of Galan, but with a different proportion giving rise to defection. As shown in Figure 6, 128 out of 1,000 runs (or 13%) of 1,000,000 generations ended in norm collapse, as opposed to 70% reported by Galan.

Replication of an entire population of non-defecting agents increases the likelihood of significant fluctuations in vengefulness over subsequent generations. For example, in one phase of a run using Galan's approach in which all agents have 0 boldness, five agents have vengefulness of 0, eleven with 1 and four with 0.8, the next generation includes eight agents with vengefulness of 0, seven agents with 1, and five agents with 0.8, simply due to the replication policy. This means that average vengefulness drops from 0.71 to 0.55 and, as boldness continues at 0, replication again makes this worse. However, replication could cause the opposite, increasing the number agents with high vengefulness over those with low vengefulness.

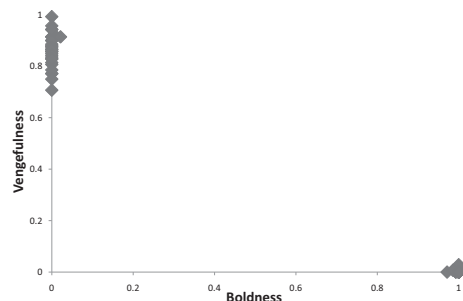
As new generations of agents with low boldness are evolved for more iterations, it becomes more likely to observe the following combination of events. First, the levels of vengefulness decrease repeatedly through a sequence of downward fluctuations until they reach very low levels. Until vengefulness levels fluctuate back in an upwards situation, this creates a temporarily fertile environment for defectors. Next, the boldness of one or a few agents increases to a high level due to mutation. This causes the bold agents to defect and be duplicated in subsequent iterations of the game. The end result of the phenomenon is an agent population where high boldness and defection are so prevalent that being vengeful leads to extinction. Thus, the game reaches a stable situation where norm collapse is ingrained in the population.

This end state is reached in a proportion of experimental runs of both Galan's and our experiments because it is more likely to reach the required preconditions when repeating the experiment for 1,000,000 generations. A much larger proportion of Galan and Izquierdo's runs end in stable norm collapse because their replication policy allows for much more significant fluctuations of vengefulness.

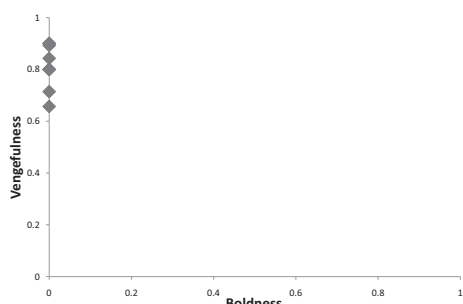
6 Mutation

As discussed above, mutation is significant in determining when norm collapse occurs. Galan and Izquierdo [7] argue that decreasing the mutation rate from 0.01 to 0.001 allows norm collapse to arise much earlier. They present an example in which a mutation rate of 0.001 allows the population to converge toward norm collapse in about 25,000 generations as opposed to 300,000 with 0.01. However, they do not explain the reasons, and do not explore the more general effect of mutation rate on norm collapse or establishment.

In seeking to consider this further, we undertook an experiment consisting of 50 runs of 400,000 generations, using the mutation rate of 0.001 suggested by Galan, with results shown in Figure 7(a). Clearly, these results support Galan's argument: 20 of 50 runs ended 76 in norm collapse. However, decreasing the mutation rate further does not have the same effect. In particular, a mutation rate of 0.0001 re-



(a) 400,000 generations and 0.001 mutation rate



(b) 400,000 generations and 0.0001 mutation rate

Figure 7: Metanorms game with different mutation rates

sulted in all runs ending in norm establishment, shown in Figure 7(b).

The relation between mutation rate and norm collapse is thus unclear. To understand this better, we performed a further series of experiments. Figure 8 illustrates the result of different experiments that consists of 10 runs each, with a range of mutation rates between 0.0001 and 0.01. As can be observed from Figure 8, the mutation rate seems to play an important role in causing norms to collapse. Decreasing the mutation rate below 0.01 has a major effect on the proportion of runs ending in norm collapse, with a peak around mutation rate values of 0.005 giving norm collapse in 80% of runs. However, decreasing the mutation rate further causes the proportion of runs ending in norm collapse to drop back (with fluctuations) until it reaches 0 with a mutation rate of 0.0001. While these results suggest a potentially interesting relationship, further work is needed to establish the exact correlation.

Nevertheless, we can say that given these results, removing mutation should avoid norm collapse. In the norms game, after the population stabilises at a low level of both vengefulness and boldness, mutation of an agent's boldness from low to high allows it to dominate, and as a result eliminate others. Figure 9 illustrates the result of a no mutation norms game that consists of 40 runs, with 1,000,000 generations each. As expected, removing mutation avoids norm collapse and leaves the population with the other two situations.

In the metanorms game, as we have seen, mutation seems to have a great effect on moving away from norm establishment. By removing mutation, we might expect to guarantee norm establishment. To corroborate this, we performed two experiments for two different durations, with results shown in Figure 10(a) for 100 generations and 100 runs and Figure 10(b) for 1,000,000 generations and 50 runs. Surprisingly, the result was not as expected. A high level of vengeful-

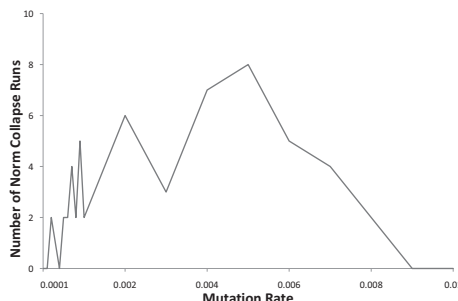


Figure 8: Metanorms game: 1m generations; 0.0001–0.01 mutation rate

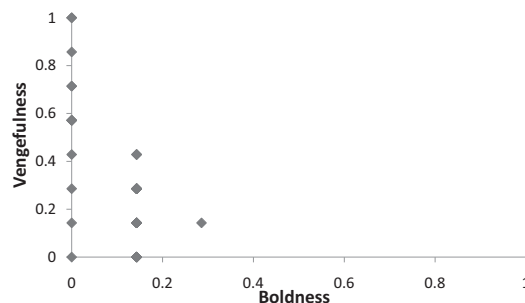


Figure 9: No mutation norms game: 1,000,000 generations

ness is maintained in almost all the runs, but a high level of boldness is also observed in some, and hence a high level of defection in the population, despite the associated punishment. This is because the final result of each run primarily depends on the initial distribution of vengefulness and boldness: individuals with high vengefulness and high boldness at the start are favoured over those with average vengefulness and low boldness. As a result, they survive and dominate the others. More importantly, the final result is determined within the very first generation so that running the experiment for longer has no impact and there is no change to the population once the levels of vengefulness and boldness stabilise.

7 Discussion and Conclusions

It is clear that Axelrod's model exhibits many interesting aspects, and relies on characteristics that provide different results with different assumptions or instantiations. We have explored several of these in relation to our experiments, and found some distinct features and, as a result, we can also provide a characterisation of the nature of norm establishment or collapse more generally.

Given the analysis through the paper, it should be clear that norm establishment lies in the region where vengefulness is high and boldness is low; similarly, norm collapse lies in the region where vengefulness is low and boldness is high. This is as used by Axelrod in his model, and underlies the aim of the initial experiments. It is illustrated graphically in Figure 11. However, we can also characterise the region where vengefulness is low and boldness is low as tending to norm collapse: this is a region of benign behaviour since boldness is low and defection is unlikely, but it is unstable since a mutation

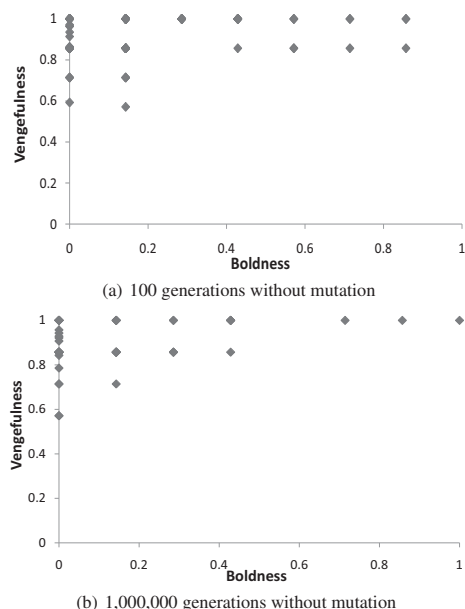


Figure 10: Metanorms game with different mutation rates

to boldness may take it higher, leaving vengefulness low, and causing norm collapse. Conversely, the region where both vengefulness and boldness are high is tending to norm establishment: it is undesirable since boldness is high and there are many defections, but these defections are likely to be punished. If vengefulness does cause punishment, then it is likely that boldness will drop, leading to norm establishment. Given this view, we can reinterpret the previous experiments. For example, it's clear that the run shown in Figure 2(a) ends in a state tending to norm collapse, the run in Figure 2(b) ends in norm collapse, and the run in Figure 2(c) ends in norm establishment, just as the run in Figure 3.

While we have analysed the impact of duration, reproduction policy and mutation rate on Axelrod's game, some of the results, such as the exact relation of mutation rate to norm collapse also require further analysis than is possible in this paper. In addition, we have only addressed some of the potentially relevant issues both due to space constraints and time in undertaking our analysis.

For example, some results in the paper suggest that the sanctioning structure does not allow a population to recover once it stabilises on high boldness and low vengefulness. In such situations, agents with average vengefulness score lower than others because of the high enforcement costs, which might be addressed by introducing a larger gap between enforcement and punishment costs. While Galan and Izquierdo [7] experiment with a set of reduced values for *meta* punishment and *meta* enforcement (but preserving the ratio), leading to norm collapse much earlier, their analysis is not extensive. Indeed, by preserving the ratio of costs but at lower values, it seems obvious that norms will collapse, because metanorms remain ineffective, as agents may pay a much higher price for enforcement than the punishment value they receive for not punishing.

Clearly, further work on the impact of sanctions on the model may be valuable in determining the exact relationship to norm collapse

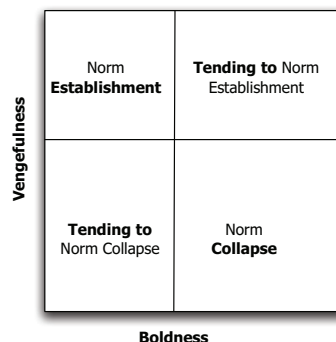


Figure 11: Characterising the vengefulness-boldness space

and establishment. Other relevant issues may include population size, temptation value (as suggested by Galan), and so on. All these are currently the subject of further analysis as our own work progresses.

REFERENCES

- [1] R. Axelrod, 'An evolutionary approach to norms', *The American Political Science Review*, **80**(4), 1095–1111, (1986).
- [2] G. Boella, L. van der Torre, and H. Verhagen, 'Introduction to normative multiagent systems', *Computational & Mathematical Organization Theory*, **12**(2-3), 71–79, (2006).
- [3] M. Boman, 'Norms as constraints on real-time autonomous agent action', in *Multi-Agent Rationality: Proceedings of the Eighth European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, volume 1237 of *Lecture Notes in Computer Science*, pp. 36–44. Springer, (1997).
- [4] F. Dignum, 'Autonomous agents with norms', *Artificial Intelligence and Law*, **7**(1), 69–79, (1999).
- [5] J. M. Epstein, 'Learning to be thoughtless: Social norms and individual computation', *Computational Economics*, **18**(1), 9–24, (2001).
- [6] F. Flentge, D. Polani, and T. Uthmann, 'Modelling the emergence of possession norms using memes', *Journal of Artificial Societies and Social Simulation*, **4**(4), (2001).
- [7] J. M. Galan and L. R. Izquierdo, 'Appearances can be deceiving: Lessons learned re-implementing Axelrod's evolutionary approach to norms', *Journal of Artificial Societies and Social Simulation*, **8**(3), (2005).
- [8] J.P. Gibbs, 'Norms: The problem of definition and classification', *The American Journal of Sociology*, **70**(5), 586–594, (1965).
- [9] M. J. Prietula and D. Conway, 'The evolution of metanorms: quis custodiet ipsos custodes?', *Computational and Mathematical Organization Theory*, **15**(3), 147–168, (2009).
- [10] B. T. R. Savarimuthu, M. Purvis, M. Purvis, and S. Cranefield, 'Social norm emergence in virtual agent societies', in *Declarative Agent Languages and Technologies VI*, eds., M. Baldoni, T. C. Son, M. B. van Riemsdijk, and M. Winikoff, volume 5397 of *Lecture Notes in Computer Science*, pp. 18–28. Springer, (2009).
- [11] D. Villatoro, S. Sen, and J. Sabater-Mir, 'Topology and memory effect on convention emergence', in *Proceedings of the 2009 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technologies*, pp. 233–240. IEEE, (2009).
- [12] F. López y López and M. Luck, 'Modelling norms for autonomous agents', in *Proceedings of The Fourth Mexican Conference on Computer Science*, eds., E. Chávez, J. Favela, M. Mejía, and A. Oliart, pp. 238–245. IEEE Computer Society, (2003).
- [13] T. Yamashita, K. Izumi, and K. Kurumatani, 'An investigation into the use of group dynamics for solving social dilemmas', in *Multi-Agent and Multi-Agent-Based Simulation*, eds., P. Davidsson, B. Logan, and K. Takadama, volume 3415 of *Lecture Notes in Artificial Intelligence*, pp. 185–194. Springer, (2005).

Modelling the social fabric for normative NPCs in MMOGs

Magnus Johansson¹, Harko Verhagen²

Abstract.

A computer game in an online setting is not something entirely new. But something that could be interesting to notice is that nowadays numerous researchers around the globe have their attention on online worlds. Most research is focused on questions such as: what is a game? Why do we play games? And how can we understand the player? Our approach is slightly different, and our aim is to make games more interesting and challenging. We believe that if we are looking at the interaction in MMOGs, the agents or “NPCs” as they are commonly referred to in games, are the most promising components of these games to improve storytelling and game-play. The problem is that game researchers often lack the skill to actually model behaviour of agents, to put it bluntly: NPCs are agents. So in order to develop the next generation of “smart” NPCs we need assistance from people skilled in deontic logic, and normative multiagent systems to develop these games further. This is an article and a proposal for future research in this field.

1 Introducing massively multiplayer online games

Massively multiplayer online games (MMOGs) are a game genre with approximately a 15 year long history, where new games have appeared on a regular basis ever since the beginning. Game designers discovered at an early point that people like to play computer games together, and with increasing computer power and Internet connection speeds, games such as “Everquest” and “World of Warcraft” have been commercial successes.

MMOGs are characterized by being persistent online worlds, with tens of thousands simultaneous players. To prevent lags in the game, players play on different servers, often called realms or shards. The fantasy genre is still one of the most common settings for MMOGs but Sci-fi settings and other settings exist. Most game worlds have a faction system where the players are divided into opposing teams in order to create a conflict.

The first step for a new player is to create a character from one of the factions by choosing race, class, gender, appearance and name. The class system of most games is balanced in such a way that certain classes are more efficient as solo playing classes, while other classes are more useful when playing in groups. Most MMOGs are designed to encourage social interaction and the class system is just one example where this becomes apparent. The different classes complement other classes with class specific skills, when players cooperate to defeat tough “Non Player Characters” (NPCs) and “MOBs” (Mobile Objects). Sometimes differentiation between what is a

MOB and what is an NPC is being made, but for the remainder of this article, we will use the term NPC for all computer - generated characters.

Depending on preferences players have different reasons for playing MMOGs, but many players try to reach the maximum skill-level in the game, often referred to as the level-cap. Some aspects of the game world are not accessible for players before they reach the level-cap. Therefore, gaining experience and leveling your character is one of the things that players strive for. In World of Warcraft the current level-cap is level 80, and to get there a player have to spend many hours (or weeks) in the game earning experience by completing quests and killing NPCs.

As mentioned above players can organize themselves in different kinds of groups. The most common types are called “Pugs” (Pick up groups), this is a small group with a short lifespan, often with a focus on completing a specific quest. Larger groups with a longer lifespan are called Guilds. Most guilds have rules and a focus on some aspect of the game. There are raiding guilds that are specialized in the specific instances of the game called raid instances. An instance is typically a cave full of monsters and treasures that are too powerful for an individual player.

As described in this brief description of MMOGs there are many opportunities for cooperation and conflicts in this type of games, making them an interesting area of research for various fields.

2 Normative multiagent systems and MMOGs

There are many ways of describing an agent, as there are to describe or define what is true about norms. We will not focus on or deliver a precise definition of agents or norms, instead we are concerned with how agents could coordinate themselves in groups, and how norms will serve as a way of organizing the agents. One definition that serves our purpose is: “social norms are always about observed behaviour” [7, p 24]. On a basic level what is important is that an action is observable and interpretable “On the level of meaningfulness human action is guided by symbols whose meaning is interpreted by actors” [5, p 119]. Intuitively norms are statements about what is appropriate behaviour in a certain context but in order to be consistent with previous work on norms, the following definition will be used for the remainder of this article: “Norms are statements about the appropriateness of an individual’s act which may result in a sanction being issued by another individual or an individual belonging to a specific class of individuals.”[10]. In addition to the definition on norms we can also note that norms are dynamic and change over time. Some will change more rapidly, while some will be fairly static. In addition to norms being flexible, the interpretation of norms is bound to be even more so.

¹ Dept. of Computer and Systems Sciences, Stockholm University, Forum 100, 16440 Kista, Sweden. Email: magnus@dsv.su.se

² Dept. of Computer and Systems Sciences, Stockholm University, Forum 100, 16440 Kista, Sweden. Email: verhagen@dsv.su.se

When discussing normative multiagent systems we can propose different traits that such a system should have and on different levels to be regarded as truly normative. We have proposed that MMOGs could be described as a normative multi agent system [10]. If we focus on the different kinds of agents in such a world, we can differentiate between two main categories: (human) player characters and non-player characters. (NPC). In MMOGs different NPCs have different levels of complexity and most of the time NPCs are just smart enough to make game play varied enough to fool a player. A technique called pathing is often used to decide how the NPCs should move in the world, according to some set of rules. For the observant player, the pathing of NPCs or the pattern in which they move can sometimes be predicted, which disrupts the illusion of the game world being inhabited by “smart” creatures. Pathing is not a problem that a normative multiagent system would solve, it is rather an example of the level of complexity the NPCs have today. Another problem with all NPCs is that they also have a predestined spawn point and spawn frequency. This means that if you kill a certain NPC, that NPC will re-spawn at fixed intervals, at a pre-determined location.

What is problematic with NPCs is that we do not see their full potential in terms of creating content, intrigue, and challenges or game-play in these games. This is of course a subjective statement but in any case much more could be done to make NPCs truly dynamic, interesting and creating a game experience where player characters actually can interact with NPCs or find them hard to outsmart just by observing their behaviour. Most of the MMOG worlds are distributed worlds accessed through a client installed on your computer, and perhaps the reason why NPCs are not allowed a higher complexity is a question of computing power and performance on the server side. If this is the case hardware development has been known to change rapidly, and this would thus be a temporary problem.

We believe that the field of Normative multiagent system research has reasons to look further at these worlds and that their contribution to the complexity of NPCs will change how we look at game worlds and human agent interaction today.

From a sociological point of view, we can potentially answer one of the classic questions about norms: when do they actually appear and at what point will a norm actually become a norm. This paper will also provide a proposal for how we can look at a normative multiagent system from a contextual “macro”-level, down to the micro-level mechanics that governs the “agent to agent” interaction.

2.1 A vision about future MMOGs

Regarding MMOGs, some say that the game actually does not get interesting until you reach what is called the “end game” content. To reach the end game you will have to play a fair amount of time being quite instrumental in your strategies to level the character as fast as possible. From a gamers point of view you can have lots of different reasons to play, and leveling your character to the level cap of that particular game perhaps is not a primary reason for playing. Still, the focus of most persistent worlds (all MMOGs are persistent worlds) is on quests and some repetitive elements where the gamer has to kill NPCs to gain experience, collecting loot and money to develop the character. These activities are sometimes called “Grinding” or

“Farming”, words that in a way communicate what a player has to endure to reach maximum level.

As mentioned above, NPCs behaviour tends to be predictable in ways that you either like or dislike. Some NPCs are easily killed due to their poor pathing and a player that has the intention of leveling will sooner or later find grinding boring, monotonous, and without variation. Some NPCs are slightly more intelligent and can, if their health reaches a certain percentage of maximum health, call for help from its nearest neighbour or run away in panic. But what if NPCs could cooperate, communicate strategies and norms in a more elaborate way in order to be unpredictable for a player?

3 NPC Types

In this section we will focus on two main types of NPCs, friendly NPCs that cooperates or helps the player, and hostile NPCs (these would traditionally be called MOBs). In most MMOGs the story behind the game revolves around some kind of conflict, since conflict is a good dramatic grip when telling a story. Given the division of NPCs into friendly and hostile NPCs all NPCs on the opposing faction in game are hostile NPCs.

3.1 Friendly NPCs

There are many NPCs that assist the player in MMOGs. Some of them are vendors where players can buy equipment or repair items. Others distribute quests for the player, where the quests most of the time involve killing hostile NPCs and collecting items that are essential in completing the quest. All these NPCs have predetermined ways of interacting with players and they are reduced to the function that they are programmed to facilitate.

A structural problem in most online worlds is that they are designed to have special places where players are meant to “socialize”. Most players will sooner or later go to a city in these games where the interaction between players is dense, and where NPCs provide different services. Our point here is that friendly NPCs could potentially have a more dynamic role in MMOGs than being vendors or quest givers. They could act as mentors tutoring new players and act as companions helping players working their way through the game to the end content. These agents will not be full surrogates for human players, but they could be a meaningful contribution to the interaction when your gaming friends are not online. If we consider making agents more responsive and interactive we should perhaps make friendly NPCs not only normative, as in that they can interpret and understand the norms of the players, but also mimic some other traits such as altruism and emotions, since altruistic actions without personal gain is something inherent in human behaviour [11]. Emotions in normative multiagent systems could be motivated through some of the sanctioning mechanisms that are used to maintain and communicate norms.

Other functions that could potentially be filled by friendly NPCs to enable more actively monitoring of norms is to enforce sanctions on for instance gold farmers, since this is a problematic behaviour that in the end leads to weakened in-game currency for all players (through mud-flation). What would the gains be from implementing friendly “law enforcement” NPCs, well one reason would be that of fairness, since Game masters (GMs) are not around in the case of a dispute between players. Disputes

that could easily be resolved most of the time with a “smart” and reactive agent that can interpret norms in a particular context.

3.2 Hostile NPCs (Mobs)

When a player is on a quest that involves exploring a cave full of dragons (hostile NPCs) the difference between encouraging instrumental play or trying to make every such quest a true adventure lies in how the NPCs behave. As mentioned above, most NPCs are fairly static and the ones that display some level of dynamic behaviour will not change their behaviour over time. This means that there is always a possibility for players to easily find a strategy in order to maximize their gain and minimize the cost of killing the dragons. If hostile NPCs could refine their tactics through cooperation and change their behaviour in response to players’ strategies, they would become harder to predict. Depending on preferences there is reason to believe that even the “achievers” from Bartle’s “Player categorisation” [1], would find NPCs with dynamic and unpredictable behaviour a much more interesting counterpart since it would demand skill and dynamic strategies to succeed in killing them.

Another scenario that would be interesting, even though it is bound to be hard to regulate the balance of a world full of such agents is, when the NPCs make up an ecosystem of their own with, e.g., predators and prey animals. If we decide to create a digital world where people will be part of some kind of balance between computer generated agents and human agents, it is fair (and much more interesting) if the NPCs actually have different societies with distinguishable norms of their own, and strategies to survive from all predators. A problem that is bound to happen in such a system is that prey animal can become an exhausted asset, since the spawn rate is the only way to regulate their occurrence that is if we are not going to implement evolutionary strategies as well.

4 How to realize our vision

To start with, we need a way of describing MMOGs on a contextual level, where we can get an understanding of the components from a macro level of the whole system, down to a detailed view of agents and how they interact. We have chosen Parsons’ “general theory of action” for the contextual level.

Next, we need a way to model how the interaction between agents works. Signalling has been used as a top-level construct in earlier attempts to implement normative multiagent systems [9]. We would however like to propose some alterations to existing systems to also incorporate mechanisms for trust and altruism between agents.

In addition to previous steps, we need to take a closer look at sanctions, rewards and the internal /external evaluation action taken by an agent.

5 Detailed description of a social system by Parsons

In guilds and clans, one of the most used strategies to make members get an idea about the norms of that particular society is to write them down on the forum or in some kind of rulebook. This is of course a way of signalling what kind of behaviour is

accepted in that context. But we believe that a normative multiagent system needs structure and differentiation into different subareas and levels to be useful, since the rules and norms of a whole MMOG, will not be made public as in the case with guilds and clans. The norm typology that we suggested [10] as a way of differentiating ways of sanctioning norm breeches should be integrated in the contextual model on a low level, in order for us to analyze the detailed interaction of agents.

Parsons elaborated a general theory of action containing three main components and a scheme for the analysis of social systems. We argue that MMOGs is a (hybrid) social system and Parsons’ scheme is particularly suitable since it is designed to describe social systems of different sizes, ranging from societies to small social groups. This is a kind of dynamic we need in order to differentiate various levels of interaction in MMOGs.

The main components are according to Parsons “Personalities”, “Social system”, and “Culture”. “Personalities” are the individual agents of the system, with personal traits, beliefs, thoughts and ideas. Relationships and the interaction between the agents create the “social system”. The “culture” in the system is the shared ideas about the social system.

These components and the interpretation of their interdependencies could potentially be what we look for in a contextual model for MMOGs as normative multiagent systems. The system should be viewed, as a whole since the interrelations between the components is what makes a working pattern.

In order to understand Parsons model we should view social interaction or social life as units, where the units are part of a larger whole (the social system). These units can be divided into classifications such as: family, friends, school, or government. Typically all “units” have a boundary, and can be seen as part of a larger whole, where there is a noticeable difference of belonging to the group, or being on the outside of it. In essence the group should have the means internally to maintain its status as a working unit, or put differently “There are two conditions any system must meet in order to survive: it must relate effectively to its environment; it must maintain its own internal integrity” [2, p 42]. These two conditions are divided into the four “functional imperatives” that follows:

1. Adaption – Any social unit has to create the means to achieve its goals as a collective. Adaption and adaptability are the means for a social group to coordinate their efforts towards goal attainment.

2. Goal attainment – Can be described as the expected output from a social unit or its goals. In order for a social system to attain its goals it has to be responding to the needs of the environment surrounding that particular system and its context.

3. Pattern maintenance and tension management – Any social unit that will maintain its cooperation must have a shared interest or shared goal. The members of the unit should ultimately be loyal to the shared goal and work towards it in order to provide “pattern maintenance”. If this proves to be hard labour that strains the resources of the group, “tension management” is imperative for the group to rest and recuperate.

4. Integration – any social unit will have relations to other social units or individuals. They will also be part of a larger whole and in order to regulate these relations, an effort towards integration is an important part of that social unit.

Figure 1 (at the end of the article) depicts some examples of different social systems and their interdependencies.

5.1 The signalling process

If we look at normative multi agent systems, we find that one of the most important aspects for the agents in the system is how they communicate the norms of the specific society. One model to describe this is called the signalling model. A model that in certain perspectives comes close to the signalling process is the work of Savarimuthu, T. et al [9] where all actions can be interpreted as a signalling process for other agents to observe in order to find potential norms. Through a procedure where candidate norms are being chosen from a list of special events that are events of specific interest, agents evaluate what will become a norm or not. What is unclear is if the world contains norms from the beginning or not, a problem that will not occur in the case of implementing normative agents in an existing MMOG, where the norms will be inherited from human agents.

5.2 Sanctions and evaluation

If we choose only to look at how sanctions are distributed when a norm has been breached, intuitively we will only find partial information of how norms are really being communicated in a society. Sanctions are probably the most obvious way of signalling what is a norm and what is not, but what happens to rewards? When an agent happens to strengthen a norm when complying with the rules and norms of a society, this should be seen as a signal that also has effect on the norms. Our assumption is that we need a normative multiagent system with the possibility to give rewards for compliant agents and altruistic behaviour, and a memory function for each agent in the system. We have briefly discussed the cost of sanctioning, the expectations from the group, and certain cases where the sanctioning is done by an agent of a specific status as mentioned in [10]. In the case of an agent that refrains from sanctioning another agent because of the cost or because of utility, a memory function of that event could prove useful. We have to keep in mind that an agent due to its status in the group might refrain from sanctioning certain norm breaches if the cost is too high for that particular agent. Furthermore, if we think about what it means for an agent to be sanctioned it is intuitive that a received punishment/sanction leads to an evaluation both internally and externally about something that we could call the agents reputation "The signalling theory suggests that any costly action can be a signal, that is, a mechanism for establishing or preserving one's reputation" [7, p 24]. It depends on if we prefer to have a focus on which norms will be communicated in our system or if we choose to look at the interaction between agents. If normative multiagent systems research wants to prove beneficial for the game industry a shift in focus towards mimicking human systems will be of importance.

5.3 Trust and responsibility – reputation

There must be a possibility for agents to remember which agents have the same beliefs about the norms of society (based on past behaviour) and which agent have a debt and should be willing to repay us in the future (incentives). If agents cannot cooperate we will lose the internal collective good [7, p 32].

Trust is something that can be hard to explain in terms of game theory, or rather it is easy to make a model but that model will

not explain why some agents act altruistically and refrain from utilitarian strategies. When we talk about trust we use Elsters' definition: "to refrain from taking precautions against an interaction partner" [3]. Different reasons for refraining from precautions exist but one of the most reasonable ideas is that the cost of taking precautions might exceed the cost of being fooled or as mentioned before that it is an inherent behaviour in humans that would be beneficial for the interaction between agents and human agents. There are a number of reasons for trustworthiness:

- past behaviour
- incentives
- signs
- signals

A way to solve this problem is to let the agent reason about reputation and confidence as proposed by [6]. Their example proved that reasoning about reputation and confidence makes it possible to see different possibilities in the choices of whom to interact with.

6 Conclusions

On a contextual level, Parsons general theory of action provides the structure needed to further reason about social systems where agents and humans interact. A system where norms change over time and where new norms are being created and communicated can help us answer at what specific point a norm becomes a norm. Norms in their infancy are statements of some kind of functional rationale that evolve into being part of a belief system that regulates how we act in a certain context.

By introducing concepts such as trust and responsibility in order to give us more detailed and observable behaviour we add to the complexity of systems inspired by the signalling theory. We argue that these additions to the signalling process are essential to create a normative multiagent system that can be integrated in MMOGs to increase the game-play, and overall impression of the AI in this game genre.

7 Future work

We have started to create a framework for further studies in this article. It is starting to be possible to make out the outlines of a normative multiagent system that closely resembles a social system with human agents and computer agents.

However some of the concepts that have been introduced above need a closer evaluation and further studies.

Sanctions and evaluation is something that we have discussed in the past [10]. Our model needs to be complemented with literature studies and a critical evaluation of an agent's external and internal evaluation of sanctions. Trust, responsibility, and reputation have been introduced on a conceptual level in this article and some indications on the direction of future studies have been made. An evaluation of the benefits from the Repage system [8] seems to be a promising next step to further develop this line of reasoning.

We have described a norm typology [10] based on the work of Gibbs [4] as an important part of our understanding of the dynamics of norms in online worlds. The norm typology will be

studied further to fill in the gaps with data from both guilds and clans in the near future.

Terminology

Farming - see grinding.
Game Master (GM) - typically an employee at the game developer that helps solving problems in the game world. Mostly they are a scarce resource and hard to get in contact with.
Grinding - when a character kills NPCs just to gain experience as fast as possible or to get some specific loot from the NPCs, as part of completing a quest. Farming on the other hand implies that a player character collects items or gold to sell it to other players, transforming in-game currency into real world currency. We will not discuss the moral or normative aspects of this behavior, other than this could be the task of NPCs to sanction players that farms money and contributes to the in-game monetary inflation (often referred to as “Mudflation”).
Level cap - the maximum level a player can reach in the game.
NPC - Non Player Character, sometimes the term “Mobs” or Mobile Objects is used. The difference between NPCs and MOBs is that every agent could be considered being a mobile object in a game world but not all agents qualify as a NPC. NPC are usually characters with a function such as presenting the player character with game content e.g. quests. MOBs are most of the time creatures that player characters can kill to get experience points and equipment/money from.
MMOG - Massively Multiplayer Online games, game worlds typically with thousands of synchronous players.

References

1. Bartle, R.: Designing virtual worlds. New Riders, 2003.
2. Cuff, E.C. Sharrock, W.W. Francis, D.W.: Perspectives in sociology, third edition. Unwin Hyman Ltd. 1990.
3. Elster, J.: Explaining Social Behaviour – More Nuts and Bolts for the Social Sciences. Cambridge University Press, 2007
4. Gibbs, J.P.: Norms: The Problem of Definition and Classification. The American Journal of Sociology 70, 586 – 594, 1965.
5. Giddens, A. Turner, J.: Social theory today. Polity press. 1987.
6. Hermoso, R. Verhagen, H.: Analyzing mechanisms using a Meta-framework for human behaviour in virtual organizations. Sixth European Workshop on Multi-Agent Systems, 2008.
7. Posner, Eric. A.: Law and social norms. Harvard University Press. 2000.
8. Sabater, I. Paolucci, M. Conte, R.: Repage: Reputation and ImAGE Among Limited Autonomous Partners.: Journal of artificial societies and social simulation vol. 9, no. 2. (2006)
9. Savarimuthu, T. et al.: Norm identification in multi-agent societies.: Normas special issue 2010 – under publication.
10. Verhagen, H. Johansson, M.: Demystifying Guilds: MMORPG-playing and norms. Proceedings of DiGRA 2009
11. Warneken, F. Tomasello, M.: The roots of human altruism.: British Journal of Psychology. 100, 455-471, 2009.

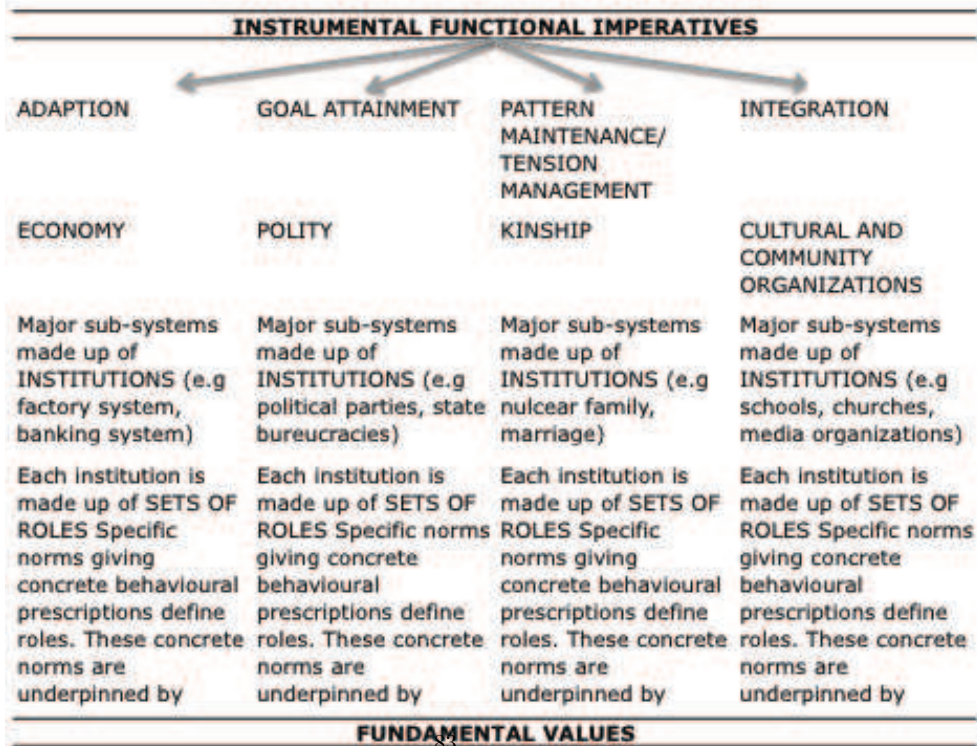


Figure 1 Parsons social system and interdependencies [2]