

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]

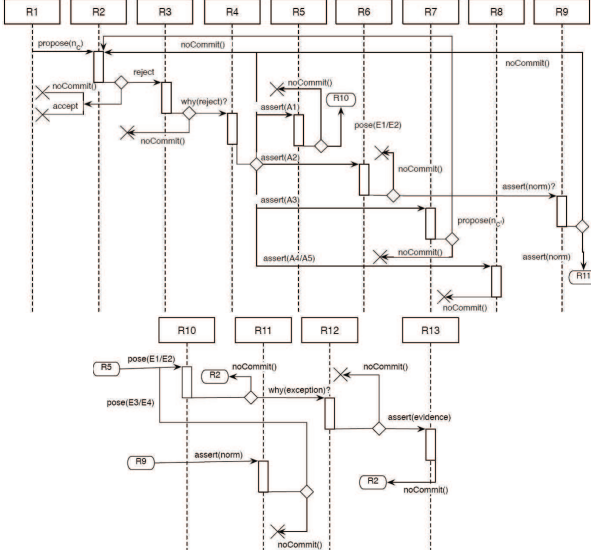


Figure 1. Sequence Diagram of the Interaction Protocol

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

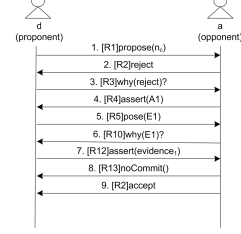


Figure 2. Case-Study Dialogue

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.

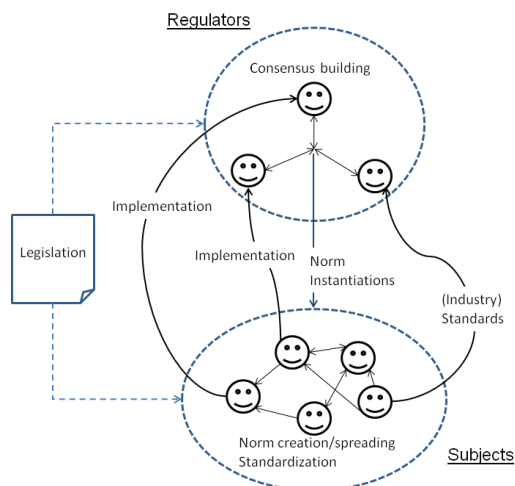


Figure 1. Levels involved in norm emergence

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 immergegence, 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 immergegence	[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 immergegence): 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.

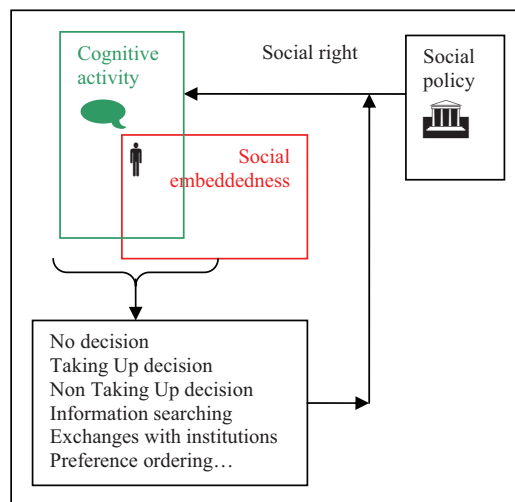


Figure 1. Individual behaviour and social rights : an overall process

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.

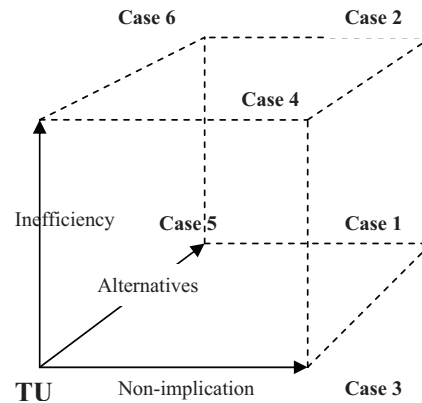


Diagram 2. Forms of non-taking up and normative positions

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.

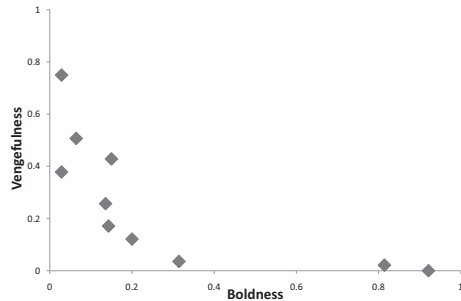


Figure 1: Norms game overall results

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.

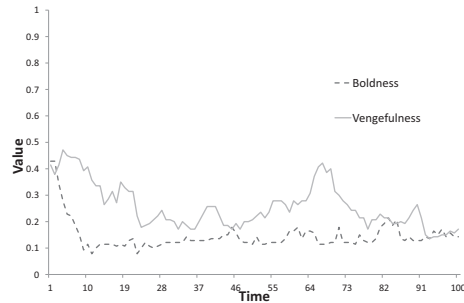
Axelrod's experiment was run five times, each with 100 generations. Two runs resulted in high average boldness and almost zero average vengefulness so there was no norm emergence at all, two other runs gave low boldness and vengefulness, but only the final run had a high level of vengefulness and very 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

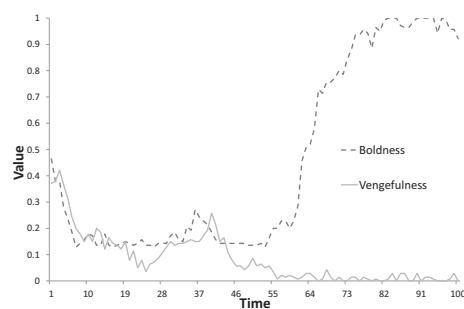
The key idea underlying Axelrod's metanorm mechanism is that some further encouragement for enforcing a norm is needed. This is accomplished by introducing a *metanorm* for punishment of those who observe a defection but do not punish the defectors. In this new metanorms game, if an agent sees a defection 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

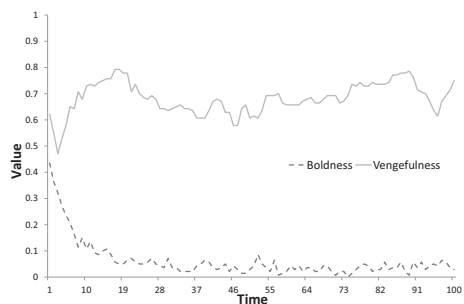
In seeking to replicate Axelrod's results [1], it becomes clear that some assumptions need to be made, about which Axelrod says nothing. First, the model does not specify how the constant population level is maintained after reproduction, when there are three possible scenarios. (i) The new population 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

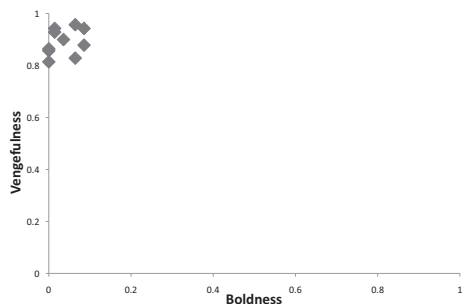


Figure 3: Metanorms game overall results

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-

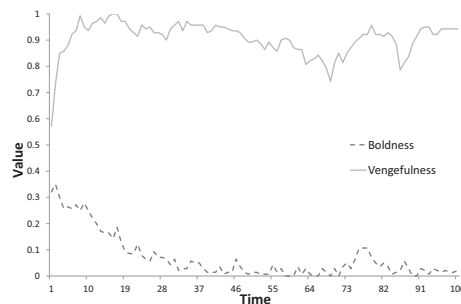


Figure 4: Metanorms game analysis

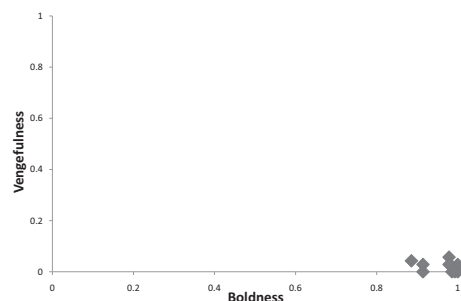


Figure 5: Norms game for 1,000,000 generations

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

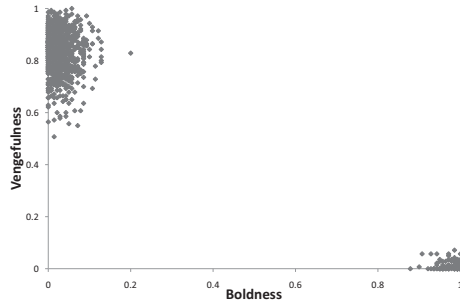


Figure 6: Metanorms game for 1,000,000 generations

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.

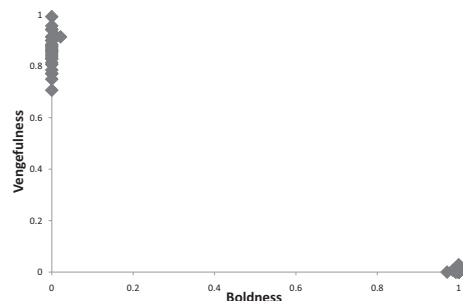
As new generations of agents with low boldness are evolved for more iterations, it becomes more likely to observe the following combination of events. First, the levels of vengefulness decrease repeatedly through a sequence of downward fluctuations until they reach very low levels. Until vengefulness 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.

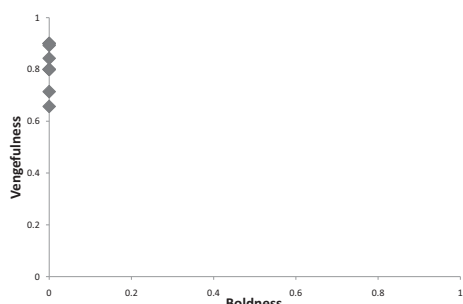
6 Mutation

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-

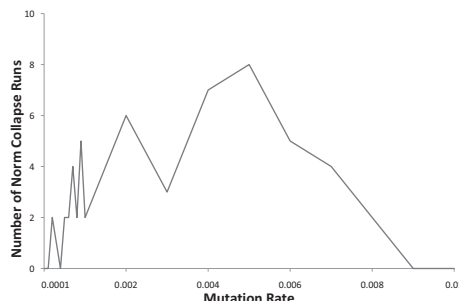


Figure 8: Metanorms game: 1m generations; 0.0001–0.01 mutation rate

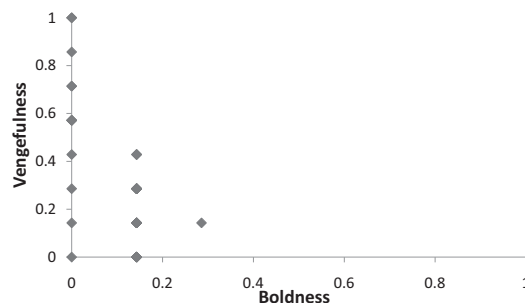


Figure 9: No mutation norms game: 1,000,000 generations

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

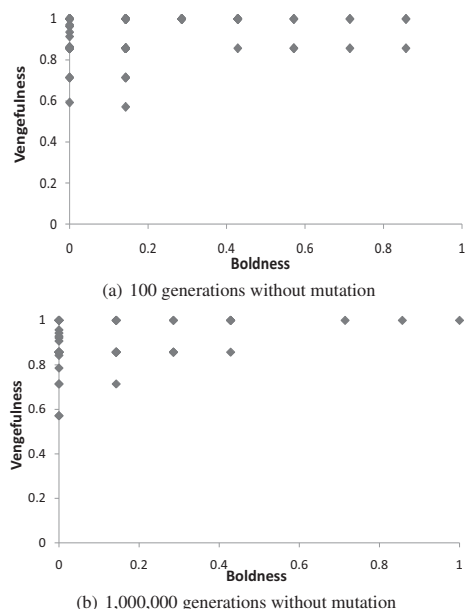


Figure 10: Metanorms game with different mutation rates

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

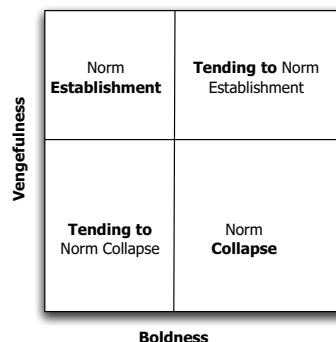


Figure 11: Characterising the vengefulness-boldness space

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.

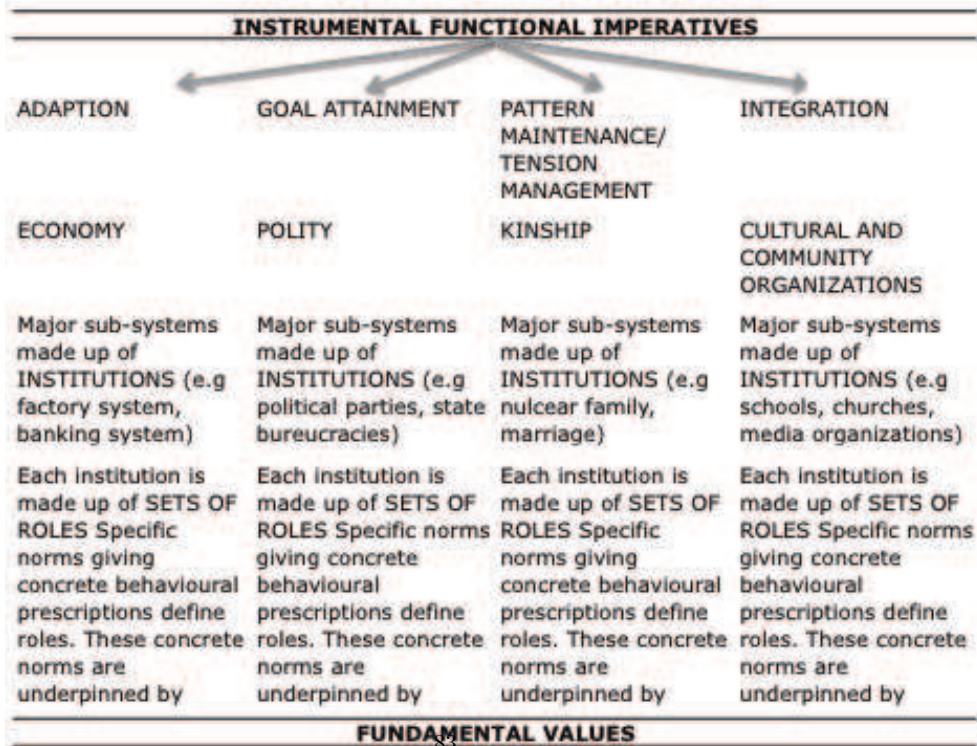


Figure 1 Parsons social system and interdependencies [2]