

# Modelling Flexible Social Commitments and their Enforcement

Philippe Pasquier, Roberto A. Flores, and Brahim Chaib-draa

Laval University, Computer Science and Software Engineering Department,  
Sainte-Foy, QC, G1K 7P4, Canada  
{pasquier, flores, chaib}@iad.ift.ulaval.ca

**Abstract.** For over a decade, agent research has shown that social commitments support the definition of open multiagent systems by capturing the responsibilities that agents contract toward one another through their communications. These systems, however, rely on the assumption that agents respect the social commitments they adopt. To overcome this limitation, in this paper we investigate the role of sanctions as elements whose enforcement fosters agents' compliance with adopted commitments. In particular, we present a model of flexible social commitments to which sanctions are attached, and where the enforcement of sanctions act as a social control mechanism for the satisfaction of commitments.

## 1 Introduction

A multi-agent system (MAS) is considered an *open* MAS if the following properties hold [1]:

1. Agents behavior and interactions cannot be predicted in advance;
2. Agents' internal architecture is not publicly known;
3. Agents do not necessarily have common goals, desires or intentions.

The first of those properties implies that the execution of open MAS is *non-deterministic*. Open societies are usually subject to unanticipated outcomes in their interactions. The second property implies that an open MAS can have members with different internal architectures; therefore, they can be *heterogeneous*. The third property implies that the members of an open society may be non-benevolent, non-cooperative or even insincere. In addition, the agents may fail to, or choose not to, conform to some of the normative aspects of the MAS in order to achieve their individual goals. In that context, providing the means and tools for the achievement of a chosen/emergent social order in such system is a challenging issue.

In the MAS literature, the reactive and proactive behavior of deliberative agents has traditionally been modelled using mental states. In systems using these types of agents, social order is naturally achieved on the assumption that agents are cooperative and collaborative, i.e., that they are sincere in following the behavior specified in the system. This raises a particularly acute problem

given that mental states are private to agents and cannot be inspected by other agents who wish to verify compliance with the specifications. This same trend has been followed in the specification of the semantics of agent communication languages, whose mentalistic definitions disqualify their use by heterogeneous agents in open systems, which are systems where agents cannot be assumed to be sincere nor to support others inspecting their internal states.

Social commitments are defined as responsibilities contracted by one agent (the debtor) toward another (the creditor), raising the expectation that the debtor will act to satisfy committed responsibilities [2]. This notion of commitment is a social one, and should not be confused with the notion of individual commitment, which emphasizes the persistence of intentions in practical reasoning, nor with the collective commitments of a group of agents. The use of social commitments removes the aforementioned drawbacks of mentalistic approaches, namely: the assumption that agents' internal architectures must conform to a particular mentalistic specification of communication semantics (embedding sincerity) and the assumption that this internal state can be inspected for verification purposes. Accordingly, social commitments can be defined independently of any agent internal architecture and decision making process<sup>1</sup>, thus supporting the development of *heterogeneous open systems* [1].

During the 1990s, social commitments were introduced as a way to capture the public aspects of communications [3]<sup>2</sup>. From then on, several research efforts have aimed at extending the use of commitments in communications as follows:

- by introducing agent communication languages (e.g., [4, 5, 6]) and models to build interaction protocols (e.g., [7, 8, 9]) using social commitments,
- by extending and complementing these approaches with theoretical [10] and practical [11] advances in the use of commitment-based agent languages.

Social commitments is a second order concept that is difficult to formalize and several social commitment-based formalizations can be found in the agents community. While some are rather restrictive and tend to consider commitments as directed obligations (e.g., using deontic logics [12]), sometimes with the possibility of unilateral de-commitments [11, 13], others are more flexible and allow the dynamic modification of commitments [3].

The notion of social commitment should be more flexible than usual obligations/prohibitions but also more rigid than permissions. More precisely, our contention is that (in the context of agent communications) social commitments are different from other normative propositional attitudes in that operations like modification, cancellation or (eventually unintended) violation should be modelled. This is important since commitments are first-class entities capturing the shared semantics of dialogues, which ground manipulations in the social layer.

<sup>1</sup> This doesn't mean that social commitments are independent of usual mental states but rather that this dependency is not part of the social commitment model *per se*.

<sup>2</sup> Although social commitments can be used for system conventions, organizational structures and agent roles, in this paper we restrict our analysis to their use in communications.

Therefore, being able to cancel or modify commitments is a key feature that allows agents to reassess the consequences of past dialogues in the context of dynamic environments. This *semantical flexibility* should not be confused with the commonly considered structural flexibility of dialogues.

This flexibility feature in social commitment models prevents their enforcement through regimentation, as is usually the case in approaches based on deontic logics<sup>3</sup>. However, given this view on flexible commitments in current ACL models, a major question remains unaddressed: what should happen in cases when agents do not respect adopted commitments? Current social commitment-based approaches are valid and useful under the assumption that commitments are generally respected. This is a strong assumption that we should inquire further. Indeed, we should describe the mechanisms supporting this functionality in open, heterogeneous systems.

In our view, the fundamental challenge is how to support the *enforcement of flexible social commitments*. This problem, which has been neglected in the modelling of recent agent communication frameworks, is addressed in this paper by introducing an ontology of social control tools for commitment-based MAS and proposing a model of social commitments compatible with our previous dialogical frameworks [8,6], which enables the introduction of the previously mentioned social control tools in open MAS.

## 2 Ontology of Sanctions and Social Control Mechanisms

Introduced in sociology as early as the end of the 19th century, the concept of social control originally denoted the capacity of a group or society to regulate itself and to secure coherency and unity in social life [15]. Social control, in this sense, relates to how social action is coordinated toward a chosen or an emergent social order. Often seen as all-encompassing, practically representing any phenomenon leading to conformity or as a broad representation of regulated mechanisms placed upon society's members, social control can be viewed as the glue holding society together [16].

Modern theories of social control focus on the strategies and techniques that help to regulate agent behavior, and lead to conformity and compliance with the rules of society (at both the macro and the micro levels). In the remainder of this section, we detail the main elements used in the enforcement of social commitments: (1) sanctions, which are considered in their general sense of incentives (the next section presents an ontology of sanctions along their different dimensions), and (2) philosophies of punishment (section 2.2), which result in punishment strategies determining the type of sanction (and its magnitude) to be applied, and explains how sanctions are assigned to social commitments.

---

<sup>3</sup> For example, due to their nature and goals, deontic logics do not even consider the possibility of violation. We refer the reader to [14] for a detailed discussion on the differences between commitments and obligations.

## 2.1 Sanctions

In this paper, we only consider *individual sanctions* and, for simplicity, leave aside other types of sanctions, such as *collective sanctions* [17] (which may be associated to teams, roles or groups of agents). In the next subsections, we go through the three main dimensions of sanctions: direction, type and style.

**Sanction Directions** – Sanctions have a specific direction. It is usually useful to consider both:

- *positive sanctions*: positive sanctions are rewards that encourage a continuation of desired behavior. For example, it is common in open systems that agents accept committing to a task only if the associated reward is worth pursuing.
- *negative sanctions*: on the other hand, negative sanctions are used to discourage norm violating behavior. For example, agents that cannot fulfill their commitments are expected to be punished.

In brief, positive sanctions are incentives to pursue a particular behavior while negative sanctions are incentives against its violation. For the sake of simplicity, we will use *sanction* to denote negative sanction in the rest of the paper, addressing positive sanctions as rewards.

**Sanction Types** – The first sanction type is *automatic sanctions*, which arises when the violator's action carries its own penalty (e.g., because it is not being coordinated with the actions of others). For example, someone who drives on the wrong side of the road has a higher than normal probability of crashing into another car. We will not consider these unintended (since no one decides that they should apply) sanctions in the scope of this paper.

Within the vast literature addressing this topic from various perspectives including economics, criminology, sociology, social psychology, AI and MAS, we encounter three broad types of non-automatic sanctions: (1) material sanctions, (2) social sanctions, and (3) psychological sanctions.

*Material sanctions* include physical sanctions like violence or repairing actions, as well as financial sanctions like fees. Material sanctions can be applied immediately at the time of occurrence or be delayed through time.

There are *social sanctions* as well. Trust, credibility and reputation are social values that could be affected by social sanctions. As pointed out in [18], social sanctions are usually the effects of some implicit informational disclosure where the violator's action conveys information about himself that he would rather not have others know. For example, that an agent violates a commitment without any explicit reason, unintentionally signals that he does not really care much about respecting the commitment, which is information that other agents could take into account when evaluating his reputation.

*Psychological sanction* types, which may be more useful in believable agents [19] and which have been used in advanced mono-agent design in mixed communities, can be important as well. Examples of psychological sanctions are guilt

(where the violator feels bad about his violation as a result of his knowledge of social norms, quite apart from external consequences), and shame (where the violator feels that his action has lowered himself either in his own eyes or in the eyes of other agents).

The time horizon of sanctions indicates whether the effects of sanctions are long-lasting or short-lived. This concept is important since some sanction types may extend through time (e.g., trust, reputation, credibility) while others may not (e.g., immediate material sanctions). Subtle and complex phenomena, like forgiveness, can require taking into account this time issue.

**Sanction Styles** – For the specific formal needs of MAS, we distinguish two sanctions styles: implicit and explicit. *Implicit sanctions* are "autonomously" and unilaterally decided by agents. The major difficulty associated with implicit sanctions is that they are not publicly known and agents have to discover whether or not they have been sanctioned (for example, by noticing that others do not communicate with them anymore). On the contrary, *explicit sanctions* are publicly known (at least among the interacting agents).

Another useful distinction can be made between *a priori* decided sanctions and *a posteriori* decided sanctions. In particular, a *posteriori* decided sanctions should be avoided in MAS, since they do not allow agents to reason about the pros and cons of respecting their commitments. That the punished agent can disagree with the sanction assigned a *posteriori* may lead to litigation.

In the remainder of the paper, we will consider only a *priori* defined explicit sanctions. Among a *priori* known explicit sanctions, we can distinguish *static*, a *priori* known, explicit, sanction systems provided to all agents at design time, and *dynamically* decided, a *priori*, explicit sanctions, which are negotiated by the agents through their communications.

## 2.2 Punishment Policies

Social control mechanisms to enforce social commitments should be designed according to a philosophy of punishment. By punishment, we mean the imposition of sanctions to satisfy open system designers' desire for retribution against wrongdoers. According to social control theorists, there are five different philosophies of punishment from which all *punishment policies* can be derived [20]: deterrence, retribution, incapacitation, rehabilitation and restoration. However, since punishment philosophies like incapacitation, rehabilitation and restoration focus on the choice of sanctions types and styles rather than on the choice of sanctions strength, we will present only the two remaining philosophies, and discuss their adequacy for open MAS:

- *Deterrence*: issued from the classical school of criminology, and supported by philosophers like Beccaria [21] and Bentham [22], deterrence is a utilitarian principle stating that the aim of sanctions is to prevent future violation. For deterrence to be effective, punishment must be swift, certain and severe. Applied to the enforcement of social commitment in MAS, it means that

commitments should be associated with heavy and explicit sanctions. This extreme position, i.e. using severe sanctions with a high prohibitive effect, tends to transform social commitments into mere obligations, losing part of the flexibility objective desired for commitments.

- *Retribution*: retribution considers that the violation should be repaired by a penalty as severe as the wrongful act.

Retribution is a practically manageable choice for open MAS. Indeed, the last decades of work in economics and law provide two basic reasons why it is best for sanctions to equal harm<sup>4</sup>. Here, we reformulate these arguments toward retribution punishment policies using MAS terminology.

The first argument concerns the *level of precautions* taken by parties, where the term "precautions" is to be interpreted generally. If sanctions are less than harm, precautions will tend to be inadequate and agents will tend to not respect adopted social commitments when it is to their advantage to do so. Symmetrically, if sanctions exceed harm, precautions will be excessive and may preclude agents committing to wanted commitments (this is the case with the deterrence punishment philosophy). For example, even if sincerely wanting to, an agent will not commit to a course of action if the sanctions attached to violation (which may occur unintentionally) are prohibitive. However, it has been shown that if sanctions equal harm, agents will have socially correct incentives to take precautions [23].

The second reason why it is desirable for sanctions to equal harm involves the agents' *level of activity*, that is, the extent to which agents participate in risky activities. An agent's level of activity affects the magnitude of expected total harm, independently of the precautions taken when engaging in an activity. For example, the more commitments an agent takes (its level of activity), the greater the possible number of accidents (violations) will occur, independently of the safety features of the agent (which affect the expected harm per commitment) [23].

It is worth noticing that concluding that damages should equal harm would require making two assumptions. The first assumption is that agents are *risk neutral*. If injurers are risk averse then the optimal level of sanctions tends to be lower than harm, because it reduces the imposition of risk on injurers and because sanctions do not need to be as high to induce injurers to behave appropriately. The second assumption is that of *strict liability*, which stipulates that injurers are definitely found liable and that no injurer cannot escape the corresponding sanctions.

### 3 Sanctions in Social Commitment-Based Approaches

In our dialogical frameworks [8, 6], social commitments are mutually established. In these approaches, agent communication is seen as the social process by which

---

<sup>4</sup> Here, harm is the violation of a particular social commitment and is at least equal to the effort that is needed to fulfill the commitment.

the social commitment layer, which captures most of the inter-agent dependencies, is manipulated. The *success conditions* of a dialogue unit (whether in a *protocol for proposal (PFP)* in [8], or in a dialogue game instance in [6]) is the social acceptance of the proposed operation on social commitments (i.e., creation, cancellation, modification or discharge) while the *satisfaction conditions* of these dialogue units are linked to the satisfaction conditions of the (eventually) resulting social commitments, which are the conditions under which a social commitment is fulfilled.

So far, we have highlighted three major points in the modelling of social commitments and their enforcement. A model of social commitments should be provided that: (1) support the semantic flexibility described earlier; (2) provide a clear, generic model for the treatment of explicit sanctions, in particular how they are attached to commitments, and the cases in which they apply; and (3) allows for a diversity of punishment policies to be supported. For example, one system could support explicit static sanctions involving only monetary transactions, while another could support agents with the ability to use explicit, dynamically negotiated, repairing actions. The next section introduces our common model of social commitments, which fulfils claims (1), (2) and (3) when explicit sanctions are used.

### 3.1 A Common Model of Social Commitment and Its Enforcement

In order to introduce sanctions within commitment-based agent communication frameworks, we first present our generic model of social commitments. Given that explicit sanctions are part of the life-cycle of commitments, the model clearly indicates the mechanisms through which sanctions are linked to commitments, and the time when they could apply. At the same time (to support claim (3)), the model is flexible enough to allow designers to control the sanctioning process.

Conceptually, social commitments are directed responsibilities contracted by an agent toward an other. We express commitments as 6-term predicates of the form:

$$C(x, y, \alpha, t, s_x, s_y)$$

where agent  $x$  is committed to agent  $y$  to satisfy the content  $\alpha$  at time  $t$ , under sanctions  $s_x$  and  $s_y$ , which specify the different sanctions that can be applied to  $x$  and  $y$  according to the states and transitions applicable to this commitment. We leave for a later section the description of the dynamics and mechanisms for associating social commitments with their corresponding sanctions. Figure 1 shows our state/transition model of social commitments, which indicates that a social commitment can be either *accepted* (denoted as  $C(x, y, \alpha, t, s_x, s_y)$ ) or *rejected* (denoted as  $\neg C(x, y, \alpha, t, s_x, s_y)$ ). It is worth noticing that a rejected commitment ( $\neg C(x, y, \alpha, \dots)$ ) is not equivalent to an accepted commitment with negative content ( $C(x, y, \neg\alpha, \dots)$ ). In particular, Figure 1 shows that a commitment can be in one of the following states:

- *inactive*: we assume a closed world hypothesis for social commitments. This means that all non-explicitly, socially-accepted commitments are rejected

(i.e., they are non-commitments), which means that they are, by default, inactive.

- *active*: a commitment is active if it has been explicitly accepted using a grounding process, and if its conditions of satisfaction (denoted as *CoS* in Figure 1) can be met.
- *violated*: this state arises when an active commitment has been violated, which is the case if it cannot be fulfilled as specified by its content (for example, if its the deadline is past) or if its content is proven to be false (in the case of propositional content). The satisfaction conditions of violated commitments cannot be fulfilled.
- *fulfilled*: a commitment is fulfilled if its content has been achieved or proven to be true (in the case of propositional content). In other words, that the satisfaction conditions of the commitment have been fulfilled<sup>5</sup>.
- *cancelled*: this state indicates that the commitment has been socially established as rejected. This state arises in the following circumstances: (1) if an active commitment is explicitly rejected through dialogue, (2) if the sanctions (possibly empty) associated with a violated commitment are applied, meaning that their compensatory effects allow cancelling the commitment, or (3) if the reward (which could also be nil) associated with the commitment is granted. This means that a cancelled commitment does not hold anymore.

Figure 1 also indicates the different transitions that could lead into these states. In order to fulfill the second requirement, we indicate the associated sanction operations for each transition:

1. *creation*: creation does not involve the application of sanctions but rather their attachment to social commitments. In particular, the negotiation of a commitment subsumes the negotiation of its sanctions. This negotiation can be complex (in the case of explicit dynamic sanction systems) or trivial (in the case of explicit static sanction systems shared among agents, where agents simply check whether or not associated sanctions match systems' conventions). This is the only transition in our model that allows the establishment of (a priori) sanctions. In particular, establishing (a posteriori) sanctions in subsequent transitions may lead to phenomena (such as litigation) that lie outside the scope of our analysis.
2. *cancellation*: cancellation deals with the rejection of an accepted active commitment, and opens the possibility of applying the corresponding sanctions. Usually, dialogical frameworks allow the hearer (the non-initiator of a dialogue) of the cancellation to decide whether or not to apply the sanctions for the cancelling agent.
3. *violation*: violation itself is not necessarily achieved through dialogue or any other mutually recognizable event. As such, the occurrence of a transition of

---

<sup>5</sup> It is worth noticing that in practice it may be possible to create an already fulfilled or violated commitment, for example, when referring to the past. However, these types of scenarios are not considered in this analysis.



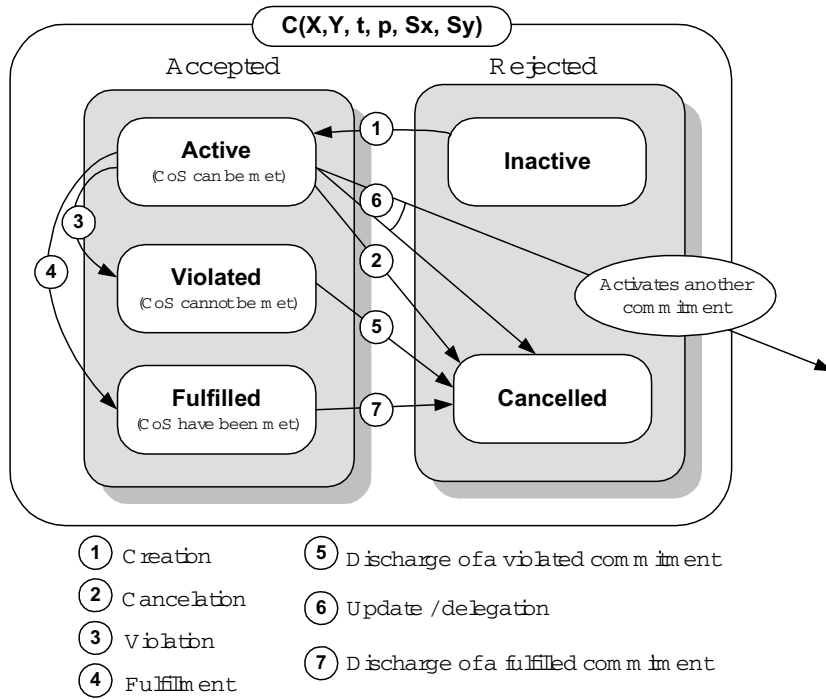


Fig. 1. The state/transition model of social commitments

this type has to be discovered and memorized for a later, socially-grounded, discharge (which is a transition described below).

4. *fulfilment*: analogously to violation, fulfilment is a transition that requires a mutually grounding event. This transition does not bring any reward but entails a socially grounded discharge where these rewards apply.
5. *discharge of a violated commitment*: the discharge of violated commitment leads to dialogues where: (1) it is socially recognized that the commitment has been violated (and that is true forever), and (2) corresponding sanctions are applied. Usually, only the debtor’s sanctions ( $s_x$ ) are applied, given its responsibility to fulfill the commitment. Furthermore, those sanctions could apply independently of the debtor’s direct involvement in the violation.
6. *update*<sup>6</sup>: updating a commitment is a double transition (as shown in Figure 1) consisting of cancelling a commitment and creating a new one (and we assume that the new one is different than the one being cancelled). Updates are socially established through dialogue and sanctions can be associated to this transition. However, updating is different from a sequencing of cancellation

<sup>6</sup> Some authors [24] consider delegation of debtor or assignment of a new creditor as particular social commitment operations. We do not, and argue that in our dialogical frameworks, delegation is a special case of updating.

and creation with respect to sanctions (which is our main reason for introducing this transition). In practice, the sanctions of updating a commitment may be less severe than those of cancelling, and (in general) it may not be unreasonable that agents could avoid sanctions altogether given satisfactory arguments. Indeed, it will be the hearer of the updating that should decide if sanctions apply or not, and to what extent.

7. *discharge of a fulfilled commitment*: discharging a fulfilled commitment leads to a dialogue in which it is socially recognized that the commitment has been fulfilled (and that is true forever) and whether the eventual reward (indicated as part of the debtor's sanctions ( $s_x$ ) as a positive sanction) is applied.

Another basic distinction (which does not appear in Figure 1) lies between *established* social commitments and *non-established* social commitments; where non-established states are inactive, violated and fulfilled (which are states that are not reached through socially grounded processes), and established states are active and cancelled (which are states that are reached through socially grounded processes). This means that transitions 1, 2, 5, 6 and 7 occur through dialogue whereas transitions 3 and 4 do not necessarily occur through social processes.

For example, in order to traverse transitions 3 and 4, we programmed our DIAGAL [6] dialogue manager (DM) as follows: each time an action is issued by an agent, the DM of that agent attempts to discover whether this action fulfills or not any of its commitments. If it does, transition 3 is applied, and a discharge dialogue game with the creditor is started to ground this fulfilment and receive potential rewards. In the same manner, all DMs check for violation (usually by observing the expiration of commitments' deadlines) and start a discharge violated commitment game if one is found. This mechanism has the property that both the grounding and the discharge of fulfilled and violated commitments are done at the moment that agents witness the occurrence of such states.

This generic state-transition model defines an operational semantic for our dialogue primitives. In particular, DIAGAL dialogue games, when successful, act as the grounding mechanisms through which transitions are realized. These dialogue games implement the sanction manipulation mechanism described above. In addition, the DIAGAL language is sound and complete with respect to its semantics, which means that all transitions are covered and nothing else<sup>7</sup>. Finally, it is worth noticing that it is possible to avoid the update and cancel transitions by not including the corresponding games in the implementation of a system, which results in a simpler but less flexible model of commitments as the one mentioned in section 1.

### 3.2 The Enforcement of the Enforcement System

One of the assumptions made in punishment-based social control mechanisms is that harming agents are always sanctioned. This assumption, which we have

---

<sup>7</sup> The same holds true for conversations under the *PPF* [4] approach.

referred to earlier as strict liability, implies that every harm is discovered and that sanctions are always applied as expected. This is a view that has been generally modelled through a regimentation hypothesis, which consists of the strict conformance of system implementations with its corresponding specifications [25]. In the case of social commitments, regimentation was assumed by implying that social commitments were systematically respected by committed agents. By arguing that this is not a realistic scenario, we shifted the focus of regimentation away from commitments and into the domain of sanctions by providing a simple mechanism to ensure efficient harm discovering (through the grounding of every transition of our model including the discharge of fulfilled and violated commitments as described in the previous section).

For the time being, we then suggest the use of software engineering constraints to ensure that the application of sanctions is treated as a strict obligation under regimentation. That is, if sanctions are not respected, they will be somehow discovered and considered as system errors. This could be done by assuming that: (1) the dialogical level rules are obligations (an agent is not allowed to digress from dialogical rules), and (2) the breach of sanctions is considered a system error (the system should be able to coerce agents into respecting the application of sanctions).

On the one hand, explicit material sanctions are easier to consider as firm obligations in a system since they are easier to verify than other sanction types. On the other hand, we assume that it should be the agents' decision whether or not to take psychological and non-material social sanctions into account. And lastly, social sanctions will usually take effect even if the agent does not realize that they exist. As such, the problem of the enforcement of sanctions thus does not have to be considered for these sanctions types.

## 4 Conclusion

In this paper we have raised the problem of the enforcement of flexible social commitment in open systems. We have introduced tools for treating this problem (section 2), namely sanctions (section 2.1) and punishment philosophies (section 2.2). We have then provided a generic social commitment model (section 3.1) that gives the operational semantic of our previous dialogical frameworks (presented in [8, 6]), supporting the enforcement of social commitments through explicit sanctions (statically specified or dynamically negotiated). Finally, we have presented the conditions under which this solution for the enforcement of social commitment does not lead to the meta-problem of the enforcement of sanctions (section 3.2). The proposed model does not make further assumptions about the agents so that the property of being able to develop open heterogeneous MAS using a flexible social commitment-based model (as introduced in section 1) is safe.

The discharge of commitments, completing their life-cycle, hasn't been fully considered in previous social commitment models. By filling this gap, we think

that our model that addresses the problem through agent communication will be helpful for further study of the long term effects of iterated dialogues consequences according to various communication pragmatics and punishment policies. Notice that the design of (domain dependent) punishment policies, linking objective (or subjective) actions values to material sanctions, is an open research issue.

## References

1. Hewitt, C.: Open information systems semantics for distributed artificial intelligence. *Artificial Intelligence* **47** (1991) 76–106
2. Castelfranchi, C.: Commitments: from individual intentions to groups and organizations. In: *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, San Francisco, CA (1995) 41–48
3. Singh, M.P.: A social semantics for agent communication languages. In Dignum, F., Greaves, M., eds.: *Issues in Agent Communication*. Springer-Verlag: Heidelberg, Germany (2000) 31–45
4. Flores, R., Kremer, R.: Bringing coherence to agent conversation. In Wooldridge, M., Ciancarini, P., Weiss, G., eds.: *Agent-Oriented Software Engineering II*. Volume 2222 of *Lecture Notes in Computer Science.*, Springer-Verlag (2001) 50–67
5. McBurney, P., Parson, S.: Agent ludens : games for agent dialogues. In: *Proceedings of the Workshop on Game Theoretic and Decision Theoretic Agents*. (2000)
6. Pasquier, P., Bergeron, M., Chaib-draa, B.: DIAGAL : a Generic ACL for Open Systems. In: *Proceedings of The Fifth International Workshop Engineering Societies in the Agents World (ESAW)*. *Lecture Notes in Artificial Intelligence (LNAI)*, Springer-Verlag (2004)
7. Fornara, N., Colombetti, M.: Defining interaction protocols using a commitment-based agent communication language. In Rosenchein, J.S., Sandholm, T., Wooldridge, M., Yokoo, M., eds.: *Proceedings of the second Autonomous Agents and Multi-Agents Systems conference (AAMAS'03)*, Melbourne, Australie, ACM Press (2003) 520–527
8. Flores, R., Kremer, R.C.: A principled modular approach to construct flexible conversation protocols. In Tawfik, A.Y., Goodwin, S.D., eds.: *Proceedings of the 17<sup>th</sup> Canadian Conference on Artificial Intelligence*. Volume 3060 of *Lecture Notes in Computer Science (LNCS).*, London, Canada, Springer-Verlag (2004)
9. Yolum, P., Singh, M.P.: Flexible protocol specification and execution: Applying event calculus planning using commitments. In Castelfranchi, C., Johnson, W., eds.: *Proceedings of the 1st International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, Bologna, Italy (2002) 527–534
10. Pasquier, P., Chaib-draa, B.: The cognitive coherence approach for agent communication pragmatics. In: *Proceedings of The Second International Conference on Autonomous Agent and Multi-Agents Sytems (AAMAS'03)*, ACM Press (2003) 544–552
11. Excelente-Toledo, C.B., Bourne, R.A., Jennings, N.R.: Reasoning about commitments and penalties for coordination between autonomous agents. In Müller, J., Andre, E., Sen, S., Frasson, C., eds.: *Proceedings of the Fifth International Conference on Autonomous Agents*, Montreal, Canada, ACM Press (2001) 131–138
12. Dignum, F., Kinny, D., Sonenberg, L.: From desires, obligations and norms to goals. *Cognitive Science Quarterly* **2** (2002) 407–430

13. Sandholm, T.W., Lesser, V.R.: Advantages of a leveled commitment contracting protocol. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence, Portland, OR (1996)
14. Vogel Carey, T.: How to confuse commitment with obligation. *The Journal of Philosophy* (1975) 276–284
15. Martindale, D.: The theory of social control. In Roucek, J., ed.: *Social Control for the 1980s: A Handbook for Order in a Democratic Society*. Greenwood Press, Westport, CT (1978) 46–58
16. Hechter, M., Opp, K.D.: Introduction. In Hechter, M., Opp, K.D., eds.: *Social Norms*, Russell Sage Foundation (2001) xi–xx
17. Levinson, D.J.: Collective sanctions. Public law research paper no. 57, NYU Law School, Ctr for Law and Business Research (2003)
18. Posner, R.A., Rasmusen, E.B.: Creating and enforcing norms, with special reference to sanctions. *International Review of Law and Economics* **19** (1999) 369–382
19. Bates, J.: The role of emotion in believable agents. *Communications of the ACM* **37** (1994) 122–125
20. Vold, G.B., Bernard, T.J., Snipes, J.B.: *Theoretical Criminology*. Fifth edition edn. Oxford University Press (2002)
21. Beccaria, C.: *On Crimes and Punishments*. New Jersey: Prentice Hall (1963)
22. Bentham, J.: *An Introduction to the Principles of Morals and Legislation*. The Athlone Press, London (1970)
23. Polinsky, M., Shavel, S.: Punitive damages. In Newman, P., ed.: *The New Palgrave Dictionary of Economics and The Law*. Volume 3. Macmillan Reference Limited, London (1998) 192–198
24. Singh, M.P.: An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law* **7** (1999) 97–113
25. Artikis, A.: Executable Specification of Open Norm-Governed Computational Systems. PhD thesis, Department of Electrical and Electronic Engineering, Imperial College London (2003)