Universitat
Autònoma
de Barcelona

Social Norms for Self-Policing Multi-agent Systems and Virtual Societies

A dissertation submitted by
Daniel Villatoro Segura
at Universitat Autònoma de Barcelona to
fulfill the degree of **PhD** in Computer Science.

Bellaterra, June 1st, 2011

Director: **Dr. Jordi Sabater-Mir**
Institut d'Investigació en Intel·ligència Artificial
Consell Superior dInvestigacions Científiques

A mis padres y a Los Beatles.

Acknowledgements

Muchas horas y dedicación están invertidas en este documento, pero afortunadamente este camino no lo he recorrido solo. Aquí sólo nombrare a algunas de las personas e instituciones que han contribuido activamente a empujarme a lo largo de este camino.

Primero de todo, tengo que agradecer a mi director de tesis, Jordi Sabater por dos motivos. Primero de todo, agradezco que hace años confiara en un africano entrevistado via Skype y por haberme dejado explorar y crecer científicamente. Gracias a él, no sólo he aprendido el método e integridad científica, sino que me ha ayudado a madurar como persona, enfrentándonos juntos a momentos adversos y alegrías insospechadas. El segundo motivo por el que le estoy agradecido es por haberme concedido esta oportunidad junto al Señor Piñol: un gran compañero y amigo, con el que sufrimos el proyecto eRep. ¿Quién diría que tantas horas descifrando y programando, nos iba a llevar a brindar (una vez más) por la ciencia en el Danubio?

Y también son protagonistas de estos brindis las siguientes personas, a las que les estoy infinitamente agradecido y que me han hecho el camino más llevadero. Meritxell, guapa!, por ser nuestro ejemplo a seguir, profesional y personalmente. Angi, por tus ánimos, fuerza y compañerismo. Sr. Nin, por el humor y los cafes. Mari Carmen, por esa sonrisa inagotable y la perseverancia ejemplar. Peter (Dani Polak), por habernos acompañado siempre. A las chicas de Administración e Inma Moros, por haberme solucionado uno y mil problemas, siempre con una sonrisa en la cara. A Tito, por ser un perdedor en la mesa de pingpong pero un buen amigo y un profesional con el cual me siento orgulloso de haber trabajado. A Elena, por tantas cervezas juntos en tantos puntos del mundo intentando sobrellevar las penas de la ciencia. A Pablo Noriega, por haberme dejado compartir muchos ratos a su lado, ilustrándome con su elegancia, y por haberme ayudado una infinidad de veces con consejos más sabios de lo que él imagina. A Pedro Meseguer, por acortar las distancias entre doctorandos y científicos con su amistad. A Juan Antonio Rodríguez, por todos sus consejos en momentos clave. Además, quiero agradecer a todos y cada uno de los miembros del IIIA, ya que para mí es un orgullo haber sido parte de esta institución, y haber compartido todos estos años con sus miembros.

I cannot forget thanking the partners from the projects I have been involved in: David de la Cruz, Jordi Brandts, Héctor Solaz, Assumpciò Vila, Jordi Estévez, Raquel Piqué, Manuela Rodríguez, Thijs, Wander, Tina, Stefan, Torsten, and the Yámanas.

I would like to thank Professor Sandip Sen, for hosting me in the University of Tulsa and sharing with me hours of work. I have learnt a lot from him and an important part of this thesis was accomplished thanks to him.

Infine, non potrò mai dimenticare i colleghi e amici di Roma. Vorrei ringraziare

Rosaria Conte per avermi accolto nel LABSS e avermi insegnato questo mestiere. Federica Mattei per i mille favori. Francesca Giardini, Gennaro di Tosto, Stefano Picascia, Francisco Grimaldo, Luca Tummolini, Walter Quattrociocchi, Mario Paolucci, Federico Cecconi e Cristiano Castelfranchi per avermi ospitato. E non ci sono parole in italiano per rigraziare Giulia Andrighetto: mi ha dato la spinta di forza per finire questa tesi. La voglio ringraziare anche per avermi trasmesso la sua passione scientifica e per essere stata un'incredibile collega e una buonissima amica. *Grazies milles* cara mia!

I want to thank the Santa Fe Institute, for choosing me to participate in their summer school, giving me the chance to meet an incredible group of scientists in one of the best research labs of the world.

Quiero agradecer también a todas esas personas que habéis confiado en mí durante estos años y que sin vosotros esta tesis no se hubiera llevado a cabo: Júlia Cot, Tomás Fuentes, Adriá Serra, Jordi López, Dani Vico, Jordi Esteller, Carles Amagat, Charly Burgmann, Pedro Martín, Jose Ramón Lozano, Alejandro Parodi, Elvira Quero, María Elías, Cecily Roche, Javier Medina, Mariola Ruiz, Gabriela Petrikovich, Maite Segu, Anna Trillo, Estibaliz Puente, Víctor Cornet, Mario Bolívar, Ingrid Dolader, Ory Shoemaker y Alfonso Rodríguez (fling).

Para finalizar no puedo dejar de agradecer a mi familia. Vuestro apoyo incondicional ha sido fundamental para alcanzar esta meta, y sólo gracias a vosotros ha sido posible. A mis tias Puri y Toñi, y a mis primas de Barcelona, por haberme hecho sentir como en casa. A mi Tata por su buen humor y todo el cariño que siempre me manda; a mi prima Inés, por no fallarme nunca; a mi cuñada Cristina por su ánimo; a mis sobrinos Pablo, Julia y Jorge por confiar en mí; a mis hermanos Diego y Fernando, mis roles a seguir y por tenerlos siempre tan cerca; y finalmente, a mi madre, por transmitirme una parte de la fuerza, el empeño y la perseverancia que la caracteriza y a todos nos fascina; y a mi padre, por haber confiado en mí y apoyarme siempre, y por las sabias palabras que una vez me enseñó y me han ayudado a acabar la "selectividad": *"Cabeza alta, paso firme y mucha mala leche"*.

Gracias a todos.

¡Por la ciencia!

Abstract

Social norms are one of the mechanisms for decentralized societies to achieve coordination amongst individuals. Such norms are conflict resolution strategies that develop from the population interactions instead of a centralized entity dictating agent protocol. One of the most important characteristics of social norms is that they are imposed by the members of the society, and they are responsible for the fulfillment and defense of these norms. By allowing agents to manage (impose, abide by and defend) social norms, societies achieve a higher degree of freedom by lacking the necessity of authorities supervising all the interactions amongst agents.

In this thesis we approach social norms as a malleable concept, understanding norms as dynamic and dependent on environmental situations and agents' goals. By integrating agents with the necessary mechanisms to handle this concept of norm, we have obtained an agent architecture able to self-police its behavior according to the social and environmental circumstances in which is located.

First of all, we have grounded the difference between conventions and essential norms from a game-theoretical perspective. This difference is essential as they favor coordination in games with different characteristics.

With respect to conventions, we have analyzed the search space of the emergence of conventions when approached with social learning. The exploration took us to discover the existence of Self-Reinforcing Structures that delay the emergence of global conventions. In order to dissolve and accelerate the emergence of conventions, we have designed socially inspired mechanisms (*rewiring* and *observation*) available to agents to use them by accessing local information. The usage of these social instruments represent a robust solution to the problem of convention emergence, specially in complex networks (like the scale-free).

On the other hand, for essential norms, we have focused on the "Emergence of Cooperation" problem, as it contains the characteristics of any essential norm scenario. In these type of games, there is a conflict between the self-interest of the individual and the group's interest, fixing the social norm a cooperative strategy. In this thesis we study different decentralized mechanisms by which cooperation emerges and it is maintained.

An initial set of experiments on Distributed Punishment lead us to discover that certain types of punishment have a stronger effect on the decision making than a pure benefit-costs calculation. Based on this result, we hypothesize that punishment (utility detriment) has a lower effect on the cooperation rates of the population with respect to sanction (utility detriment and normative elicitation). This hypothesis has been integrated into the developed agent architecture (EMIL-I-A). We validate the hypothesis by performing experiments with human subjects, and observing that behaves accordingly to human subjects in similar scenarios (that represent the emergence of cooperation).

We have exploited this architecture proving its efficiency in different in-silico scenarios, varying a number of important parameters which are unfeasible to reproduce in laboratory experiments with human subjects (because of economic and time resources).

Finally, we have implemented an *Internalization* module, which allows agents to reduce their computation costs by linking compliance with their own goals. Internalization is the process by which an agent abides by the norms without taking into consideration the punishments/sanctions associated to defection. We have shown with agent based simulation how the Internalization has been implemented into the EMIL-I-A architecture, obtaining an efficient performance of our agents without sacrificing adaptation skills.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Social norms have been claimed to be one of the research topics in multi-agent systems to be investigated [Luck et al., 2005]. Social norms are part of our everyday life, and they have been of interest in several areas of research [Elster, 1989]. Social norms help people self-organizing in many situations, specially in open-environments where having an authority representative is not feasible. On the contrary to institutional rules, the responsibility to enforce social norms is not the task of a central authority but a task of each member of the society. From the book of Bicchieri [Bicchieri, 2006], the following definition of social norms is extracted:

> *"The social norms I am talking about are not the formal, prescriptive or proscriptive rules designed, imposed, and enforced by an exogeneous authority through the administration of selective incentives. I rather discuss informal norms that emerge through the decentralized interaction of agents within a collective and are not imposed or designed by an authority"*.

Although this definition does not pretend to be a formal definition, it gives us the notion that social norms are used in human societies as a mechanism to improve the behavior of the individuals in those societies without relying on a centralized and omnipresent authority. In recent years, the use of these kinds of norms has been considered also as a mechanism to regulate virtual societies and specifically societies formed by artificial agents ([Saam and Harrer, 1999, Shoham and Tennenholtz, 1992, Grizard et al., 2006]), and in some cases, mixed populations (agents and humans).

The study of social norms within multi-agent systems has been framed inside the Computational Social Science community, which is *"in charge of collecting and analyzing data at scale that may reveal patterns of individual and group behaviours"* [Lazer et al., 2009]. The combination of both techniques (computer science and social science) makes special sense when we deal with multi-agent systems, as agents are social entities. The exploitation of the knowledge provided by the social sciences is becoming more useful with the integration of humans and agents (E.g. [Brito et al., 2009]).

Latterly, and due to the interest of the multi-agent community on social norms, the NorMAS movement has been founded [Nor, 2008]. This movement gathers an interdisciplinary community (mainly formed by computer scientists, social scientists, psychologists, and economists) around the topic of the norms in multi-agent systems. From one of the workshops organized by this community, the following definition of normative multi-agent system was agreed by consensus:

> *"A normative multi-agent system is a multi-agent system organized by means of mechanisms to represent, communicate, distribute, detect, create, modify, and enforce norms, and mechanisms to deliberate about norms and detect norm violation and fulfillment"* [Boella et al., 2008].

The online emergence of social norms (how they are created at first time) in a decentralized way is one of the problems in the community that needs to be solved and some authors are trying to approach. As we are focusing on decentralized open virtual societies, different norms might emerge on different areas of the social topology. However, social norms are by definition ([Coleman, 1998]) more socially efficient when the whole population abides by them. We are interested in studying the spreading and acceptance of social norms, what Axelrod [Axelrod, 1986] calls *norm support*. Our understanding of norm support deals with the problem of which norm is established as the dominant when more than one norm exists for the same situation, or in those situations where the agents' self-interest is against the interest of the society, and therefore violating the norms is the strategy that provides the highest payoff.
There are basically two well-known approaches to study this emergence: the game-theoretical and the cognitive. The former studies the process by which agents calculate the cost-benefit ratio of the possible actions taking into consideration the existence of a norm [Villatoro et al., 2010] and the strategic behavior of the other agents involved in the interactions. On the other hand, the latter studies the emergence from the point of view of the beliefs, desire and intentions of the agents and their relation to normative beliefs [Andrighetto et al., 2010b].

After reviewing the literature on both research approaches, we agree with Young [Young, 2008] that there are three mechanisms by which norms emerge, that generalize all the factors previously treated in the literature:

- Pure Coordination: These are "social" phenomena, because they are held in place by shared expectations about the appropriate solution to a given coordination problem, but there is no need for social enforcement.
- Threat of social disapproval or punishment for norm violations.
- Internalization of norms of proper conduct.

The first mechanisms have been widely studied under the topic of convention emergence [Shoham and Tennenholtz, 1997a, Kittock, 1993, Delgado et al., 2003, Mukherjee et al., 2007, Sen and Airiau, 2007, Walker and Wooldridge, 1995] (e.g. the typical example of these kind of mechanisms are which side of the road should cars drive: both agents are benefited from following the coordination norm, otherwise, their utility is drastically reduced).

However, we have observed a general practice in that area of research which consists in establishing the system as converged when 90% of the population shares the same convention. Because of the definition of convention (a norm that establishes a focal point amongst the possibles to promote coordination when shared by all participants), we cannot consider 90% as an acceptable convergence rate and we perform an exhaustive study on that area. We focus on aspects like the strategy update rule used by agents or the topology that fixes their interactions to study full convergence (100%) in the system. We will observe how certain strategy update rules and topologies promote the existence of subconventions, that delay the emergence of a global convention. We propose socially-inspired methods to dissolve subconventions and allow the society to reach the desired full-convergence.

On the other hand, the second and third points are tightly related amongst them. Sanctions serving as a mechanism for norm emergence has been widely studied by social scientist [Coleman, 1998], psychologists [Bandura, 1976] and economists [Fehr and Gachter, 2002]. In this thesis we will study how different punishment technologies can reinforce the process of norm emergence, allowing the system to self-policy. We initially envisioned to donate agents with the necessary mechanisms to achieve *Multilateral Costly Punishment* [Posner and Rasmusen, 1999]. In order to be realistic with the parameter setting for the Multilateral Costly Punishment, we performed experimental economics experiments that proved us the existence of a stronger force behind the punishment: the normative message affecting a normative decision making. With the light thrown in those experiments, we decided to develop a cognitive agent architecture that not only performs a benefit-cost calculation with respect to the existence of norms, but it also considers the fact of the norm's existence as determinant in the decision making. In this part we will also prove the determinant difference between punishment and sanction, which is an important element for our developed architecture, and a state-of-the-art discovery in the behavioural economics literature. As we said previously, the third point affirmed by Young for norms to emerge (Internalization) is in a direct relationship with our research on punishment technologies. In our work [Andrighetto et al., 2007, Conte, 2009, Andrighetto et al., 2010b], we intend *Internalization* as the process where an agent becomes compliant with norms not because the fear of potential punishment (and utility detriment) but just because the mere existence of norms and the willing to abide by them. We will observe how this process ensures a more robust society against norm-violations, without sacrificing adaptation skills.

1.2 Contribution

This thesis contributes to the field of self-organized and normative multi-agent systems in three lines:

First - Exhaustive Exploration of the Dimensions in the Convention Emergence Problem

We present an experimental framework to analyze the dynamics in the process of

convention emergence, observing the effects of the different parameters configuring the model. Specially, we focus on the problem of the emergence and dissolution of subconventions. We have detected how these subconventions emerge mainly for two reasons: the strategy update rule promoting concordance with previous history (and culture maintenance) and topological conditions promoting endogamy. The conducted research took us to discover the existence of Self-Reinforcing Structures, mainly in Scale-Free Networks, producing subconventions to remain metastable. We propose socially-inspired mechanisms that dissolve these Self-Reinforcing Structures, allowing the society to achieve full convergence.

Second - Experimental Platform for Regulated Hybrid Experiments

In order to obtain experimental results about certain punishment technologies, a platform allowing to test the behavior of humans under certain conditions (achieved by the partners interactions) was necessary. Another requirement for the platform was a restricted behavioural set of available actions to agents and a pre-established interaction protocol. Because of the previous reasons, the paradigm of Electronic Institutions seemed the most adequate.

We have developed a friendly user web-interface for human subjects to interact with other subjects inside an Electronic Institution. With the usage of this platform we allow humans to interact from remote locations through a simple web interface. On the other hand, the electronic institution paradigm allows us, as experiment designers, to implement agents that perform in an specific way, locating human subjects in specific interesting experimental conditions.

This platform has been essential to perform behavioral economics experiments, introducing agents into them, and allowing subjects to participate remotely, advancing the state-of-the-art platforms (like Z-tree [Fischbacher, 2006]).

Third - EMIL-I-A Architecture for Self-Regulating Societies

We have developed a BDI agent architecture which distinguishes at a cognitive level the difference between *punishment* and *sanction*. Experimental results with human subjects confirmed the difference between punishment (utility detriment) and sanction (utility detriment plus norm elicitation) in the decision making of human subjects. This difference is incorporated precisely in our agent architecture (EMIL-I-A), producing effects on the decision making of the agents at several levels. Moreover, we test the performance of our EMIL-I-As in the same experimental conditions than with humans. These experiments show that EMIL-I-As produce the same dynamics than those showed by humans, obtaining an state-of-the-art agent architecture with punishment capabilities able to interact (accordingly) with humans.

Finally, we implement an *Internalization* mechanism that allows norm compliance to happen without the external enforcement of punishment, proving the adaptive capabilities of EMIL-I-A while saving in cognitive load.

1.3 Overview and Structure of the Thesis

This thesis is structured in nine chapters:
Chapter 2: We present a game-theoretical difference between conventions and essential norms. This difference is basic to structure the rest of the thesis. Then we make an analysis of the state of the art in the MAS literature in both fields. With respect to conventions, we review the different works performed on the emergence of conventions and how they treated different aspects of such process. On the other hand, we review the different works and mechanisms analyzed for the emergence of cooperation. We focus on those that are applicable for self-organizing societies, paying special attention at MAS with punishment technologies incorporated.
Chapter 3: In this chapter we present the general framework for the convention emergence problem. This framework allows us to study the different dimensions affecting the convention emergence, focusing in the achievement of full convergence. A detailed analysis of the search space of parameters allows us to discover that full convergence was not previously achieved because of certain strategy update rules and certain topological configurations promoting subconventions.
Chapter 4: In order to dissolve the identified subconventions, we propose socially-inspired techniques that also accelerate the process of convention emergence. With the usage of these instruments we discover the existence of Self-Reinforcing Structures, and propose a combined social instrument for dissolving the subconventions created in those areas of the topology.
Chapter 5: In this chapter we propose distributed punishment as a punishment technology that can produce norm emergence in situations represented by common good games. Experimental results with human subjects give us the intuition that other types of punishment may exist and affect differently to the decision making of the agents.
Chapter 6: We present the EMIL-I-A architecture, conceived to interpret the difference between punishment and sanction at a cognitive level. This architecture is constructed assuming the existence of normative beliefs and goals, which are orchestrated with the norm salience, representing the degree of activation in the social environment of the different recognized norms.
Chapter 7: In this chapter we prove the correctness of the EMIL-I-A architecture by simulating the same experiments where human subjects participated, obtaining similar results. Moreover, we exploit our agent architecture in different experimental conditions, understanding the dynamics of our agents when interacting in different environmental situations. Finally we introduce a Dynamic Adaptation Heuristic of the cost of punishment in order to achieve a more efficient sanction, profiting from its cognitive load, and maintaining high cooperation rates.
Chapter 8: This chapter discusses a cognitive mechanism that allows agents to internalize norms, and therefore, comply with them without external enforcement. By means of simulation in a P2P inspired scenario, we will observe how internalization allows agents to efficiently respond to their peer needs, adapting to environmental conditions and dynamically changing the normative scheme.
Chapter 9: We conclude our research of normative self-organized systems and sketch future lines of research to exploit the work presented in this thesis.

1.4 Related Publications

The following publications are a direct consequence of the development of the thesis.

- G. Andrighetto, D. Villatoro, F. Cecconi, R. Conte. Simulazione ad Agenti e Teoria della Cooperazione. Il Ruolo della Sanzione. Sistemi Intelligenti (forthcoming)
- G. Andrighetto, D. Villatoro, R. Conte. Norm Dynamics within agents. In B. Edmonds Dynamic View of Norms. Cambridge University Press (forthcoming)
- D. Villatoro, J. Sabater-Mir and S. Sen. Social Instruments for Robust Convention Emergence. Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011). (In press).
- D. Villatoro, G. Andrighetto, R. Conte and J. Sabater-Mir. Dynamic Sanctioning for Robust and Cost-Efficient Norm Compliance. Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011). (In press).
- G. Andrighetto, D. Villatoro. Beyond the Carrot and Stick Approach to Enforcement: An Agent-Based Model. Proceedings of the European Conference on Cognitive Science, New Bulgarian University, Sofia, 21-24 May 2011. Cognitive Science Society.
- T. Balke and D. Villatoro. Operationalization of the Sanctioning Process in Hedonic Artificial Societies. Proceedings of the AAMAS Workshop on Coordination, Organizations, Institutions and Norms (COIN @ AAMAS 2011).
- D. Villatoro, J. Sabater-Mir and S. Sen. Social Instrument for Convention Emergence. Proceedings of 10th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS 2011).
- D. Villatoro, S. Sen and J. Sabater-Mir. Exploring the dimensions of Convention Emergence in Multi-agent Systems. Advances in Complex Systems (ACS) Volume No.14, Issue No. 2 pp 201-227. (2011).
- G. Andrighetto, D. Villatoro, R. Conte and J. Sabater Mir. Simulating the relative effects of punishment and sanction in the achievement of cooperation. Proceedings of the Eighth European Workshop on Multi-Agent Systems (EUMAS10).
- R. Conte, G. Andrighetto, D. Villatoro. From Norm Adoption to Norm Internalization. (DEON 2010).
- G. Andrighetto, D. Villatoro, R. Conte. Norm internalization in artificial societies. AI Communications. Vol No.23, Issue No.4 pp.325-339. (2010)
- D. Villatoro, S. Sen and J. Sabater-Mir. Of Social Norms and Sanctioning: A Game Theoretical Overview. International Journal of Agent Technologies and Systems, Vol. 2, Issue 1, pp 1-15. (2010).

- A. Vila-Mitjá, J. Estévez, D. Villatoro and J. Sabater-Mir. Archaeological Materiality of Social Inequality Among Hunter-Gatherer Societies. Archaeological Invisibility and Forgotten Knowledge: Conference Proceedings, Lódz, Poland, 5th7th September 2007. Archaeopress. 2010

- D. Villatoro, S. Sen and J. Sabater-Mir. Topology and memory effect on convention emergence. Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology. (IAT 2009).

- D. Villatoro, N. Malone and S. Sen. Effects of interaction history and network topology on rate of convention emergence. Proceedings of 3rd International Workshop on Emergent Intelligence on Networked Agents (WEIN'09 @ AAMAS).

- D. Villatoro and J. Sabater-Mir. Dynamics in the Normative Group Recognition Process. Proceedings of IEEE Congress on Evolutionary Computation (IEEE CEC 2009).

- I. Brito, I. Pinyol, D. Villatoro, J. Sabater-Mir. HIHEREI: Human Interaction within Hybrid Environments. Proceedings of 8th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS 2009) *(Best Student Demo Award)*.

- D. Villatoro and J. Sabater-Mir. Group Recognition through Social Norms. Proceedings of 8th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS 2009).

- D. Villatoro and J. Sabater-Mir and S. Sen. Interaction, observance or both? Effects on convention emergence. Proceedings of Dotzè Congrés Internacional de lAssociaciÓ Catalana d'Intelligència Artificial (CCIA 2009).

- D. Villatoro and J. Sabater-Mir. Towards the Group Formation through Social Norms. Proceedings of the Sixth European Workshop on Multi-Agent Systems (EUMAS08).

- D. Villatoro and J. Sabater-Mir. Towards Social Norm. Proceedings of the Eleventh International Congress of the Catalan Artificial Intelligence Association (CCIA08).

- D. Villatoro and J. Sabater-Mir. Mechanisms for Social Norms Support in Virtual Societies. Proceedings of the Fifth Conference of the European Social Simulation Association (ESSA08).

- D. Villatoro and J. Sabater-Mir. Categorizing Social Norms in a Simulated Resource Gathering Society. Proceedings of the AAAI Workshop on Coordination, Organizations, Institutions and Norms (COIN @ AAAI08).

- D. Villatoro and J. Sabater-Mir. Norm Selection Through Simulation in a Resource-Gathering Society Proceedings of 21st European Simulation and Modelling Conference (ESM07).

- J. Sabater-Mir, I. Pinyol, D. Villatoro and G. Cuni. Towards Hybrid Experiments on reputation mechanisms: BDI Agents and Humans in Electronic Institutions. Proceedings of 12th Conference of the Spanish Association for Artificial Intelligence (CAEPIA07).

Chapter 2

State of the Art

Descriptions of tasks like greeting another person, dressing, driving, etc. are often accompanied by the phrase "in a proper way". In human societies, the "proper way" to fulfill these interaction protocols is specified by social norms. A number of tasks that require some kind of interaction with other agents might require agents to follow specified guidelines to successfully complete these tasks. Social norms can facilitate such agent interactions and enable agents to complete these tasks efficiently. Such social norms can emerge and spread among the society until they are widely accepted and adopted. Therefore, we can view social norms as key elements that enable coordination and self-organization in our everyday life.

Not all the social norms, however, deal with the same kind of interaction scenarios. We observe that social norms like greeting (shaking hands, kissing, leaning towards each other, or a simple "hi!") pertain to different situations compared to, for example, the social norm of recycling. We also observe that social norms, though referring to the same concept, are defined using different terms in the literature, e.g. norms, social laws, conventions, social norms.

In addition to the different types of norms, and the wide variety of terms used to define this social instrument, the study of social norms is made more challenging by the heterogeneous perspective on this issue and how it is viewed in diverse research areas such as economics, social sciences or multi-agent systems. We believe that though these areas have interesting theories and practices to contribute to social norms literature and can benefit from prudent adaptations and applications of social norms, not enough attention and effort has been expended on this potentially effective social coordination mechanism by the corresponding research groups.

The primary goal of this chapter is to develop a characterization of social norms into two primary groups: coordination norms and essential norms. This division is performed from a game-theoretical point of view with the goal of understanding the process of norm emergence.

Before proceeding, we need to characterize the type of environment where the contribution of this thesis is though to be applied in. Technically speaking we assume the existence of a MAS environment –observable as an objective entity– with fixed ontology, dynamic state, regimented with constitutive conventions and where the par-

ticipation of capable and entitled agents is permitted.[1] The decision-making processes of agents may be not under the control of the MAS. Agents may respond to their own motivations and be owned by different owners. Agents may enter or leave the MAS at will, as long as they are capable and entitled to be in the MAS.

The type of norms treated in this work are those norms created through the interaction of agents and that are not imposed by any central authority. These emerging norms possess interesting dynamics that have to be studied to achieve real self-regulated societies.

2.1 Categorizing Norms

In the multi-agent systems literature, several terms have been used for the same concept (convention, social norm, social law) although referring to a similar term. However, those terms have been used as synonyms, and referring to a concept that was not specifically described. Considering norms as regularities in the behavior of agents, Coleman [Coleman, 1998] defines two main types of norms: conventions and essential norms. In this section we ground on a game-theoretical paradigm to clearly define conventions and essential norms, as mechanisms that solve coordination problems although in games with different dynamics.

2.1.1 Conventional Norms

Conventions (or *conventional norms*) are a type of norms that naturally emerge, with no threat of punishment. *Conventional norms* solve coordination problems (a representative payoff matrix of these type of problems can be seen in Figure 2.1), where there exist no conflict between the individual and the collective interests, as what is desired is that everyone behaves in the same way, without any major difference on which action agents are coordinated.

Following Coleman's theory, the selection of the focal action[2] in such norms is arbitrary, although most of the times is strongly biased by cultural patterns. One clear example of these kind of norms is the selection of which side of the road to drive on.

[1] That is (i) There exists a *world* that may be populated by (ii) *agents* whose activity may involve other agents and possibly other entities –like speed limits, traffic lights, or fines. (iii) At any point in time, that world is in a *state* which is represented by a set of variables whose values may change due to the actions of agents or to their own dynamics. (iv) Out of the many conceivable actions that, in principle, agents may attempt, the only ones that are deemed *admissible* in that world may have any effect in the world. (v) All actions are subject to preconditions (on the state of the world) that make them feasible and post-conditions that determine their intended effect in the world.

[2] The term *focal action* is directly borrowed from game theory where it exists the concept of *focal point*. A focal point is defined as a solution that players will tend to use in the absence of communication, because it seems natural, special or relevant to them. For example, imagine that you and your partner are visiting Paris. It is the first time for both of you in that city. Unfortunately, you are not in the same hotel and you have no means to communicate with each other, although you know that you have to meet each other at a certain time in a public place. You can choose between all the public places in Paris. Using a common sense reasoning, you might choose to go to the Eiffel Tower, the Pyramid of the Louvre Museum, or the Arc de Triomphe. Those would be focal points in the decision game. Consequently the focal action will be the action taken by the agents in the absence of communication.

Player 1 \ Player 2	Action A	Action B
Action A	+1 +1	-1 -1
Action B	-1 -1	+1 +1

Figure 2.1: Coordination Game

This definition is in concordance with that of Young [Young, 2007]: *"A convention is a pattern of behavior that is customary, expected, and self-enforcing. Everyone conforms, everyone expects others to conform, and everyone wants to conform given that everyone else conforms."*.

Before any agent has a preference for any convention, all the possible conventions are equally beneficial at the utilitarian point of view, and therefore, selecting which one to use is initially chosen randomly. Its spread along the society (and how it is affected by *social learning*) has been widely studied as we will see in Section 2.2.

Convention emergence is a process that starts with a population of agents with no initial preferences for any of the existing conventions, and they have to agree on a common one (because a convention, by definition, is only useful when all agents share the same convention); once all agents have agreed on the same convention, the system has converged. Pertinent examples of conventions in Multi-agent Systems are communication protocol (the language an individual chooses for communication), behavioral patterns (coordination protocols like the side of the road where to drive or how to greet when introduced to other people), or the selection of the task to be executed in a multi-task scenario.

2.1.2 Essential Norms

Essential norms are norms that solve or ease collective action problems, where there is a conflict between the individual and the collective interests, making this the main difference with respect to *conventions*. These norms are called *essential norms*. Following the definition of focal actions, in these kind of norms the focal action is not chosen randomly because the targets' interests lie in the direction of action opposing observance of the norm, and the beneficiaries' interests lie in the direction of action favoring observance of the norm. *Essential norms* help address situations where the individuals are tempted not to contribute to the common good. These problems are commonly known in the literature as *collective action problems*. Heckathorn [Heckathorn, 1996] claims that any collective action system has two characteristics:

1. Non-excludable: Excluding anyone from consumption of the common good is impractical. For example, scabs benefit from higher wages won through strikes.

2. Jointness of supply: the degree to which the good costs the same to produce regardless of the number of people who consume it. A radio broadcast has very

Player 1 \ Player 2	Cooperate	Defect
Cooperate	3 3	5 0
Defect	0 5	1 1

Figure 2.2: Prisoner's Dilemma

high jointness of supply because the costs of production have very little, if any, proportion to the number of people who consume the broadcast. Manufactured products, in contrast, typically have low jointness of supply because the cost of production increases with the number of consumers, though economies of scale may make the increase in cost less than directly proportional to the increase in consumers [Barros, 2007].

Referring to the first of these characteristics of collective action systems, in the case of the collective actions, conflicts appear when an individual is tempted not to contribute to the common pool, leading to an incomplete collective action. The fact that an individual behavior affects the welfare of the group is enough for the group to acquire the right to control individual behavior. Social norms are applied to control the individual behaviour in these kind of problems.

It has been proven [Hardin, 1971] that a *Collective Action Game* is equivalent to the *Prisoner's Dilemma*[3]. In a game that follows the structure of a Prisoner's Dilemma (like the one shown in Figure 2.2) the individually rational strategy is to *Defect*, no matter what the other player decides. The equilibrium state of those decisions is suboptimal in the sense of Pareto, as there exists a different focal point (the mutually cooperative outcome) where the outcome when both agents are coordinated is preferred by both players. A social norm prescribing cooperation would help agents in obtaining a socially-efficient behavior.

However, conflict between the individual's interest and the collective's interest can occur in other situations than that captured by the *Prisoner's Dilemma* game. Such conflicts can also have the form of the *Collective Action Game*, as shown in Figure 2.3. Only the payoffs for the individuals are shown, and they are calculated considering that V is the value of the collective good, c is the cost for the individual to contribute, and p is the proportion of the total value that an individual can produce by itself.

[3]The Prisoner's Dilemma states: Two suspects are arrested by the police. The police have insufficient evidence for a conviction, and, having separated both prisoners, visit each of them to offer the same deal. If one testifies (defects) for the prosecution against the other and the other remains silent (cooperates), the betrayer goes free and the silent accomplice receives the full 5-year sentence. If both remain silent, both prisoners are sentenced to only six months in jail for a minor charge. If each betrays the other, each receives a five-year sentence. Each prisoner must choose to betray the other or to remain silent. Each one is assured that the other would not know about the betrayal before the end of the investigation. How should the prisoners act?

Individual \ Collective	Cooperate	Defect
Cooperate	V-c	pV-c
Defect	V(1-p)	0

Figure 2.3: Collective Action Game

In this thesis we will focus on the subgroup of essential norms that can be characterized by the Prisoner's Dilemma, and we refer the reader to [Villatoro et al., 2010] for a further classification of Collective Action Games.

2.2 Convention Emergence

In real self-organized systems (and in most of the work in the literature) convention emergence can be achieved through *social learning*. The social learning process is a complicated task: agents are learning from the other agents whom they interact with, which at the same time are also learning from those interactions. Therefore, the structure of the social network that restricts the interactions between agents is essential for the emergence of conventions learnt through social learning.

This section will be useful to help us understand the roots of this area of research, and how it evolved along the years, leaving an important gap unexplored, that we will cover in this thesis.

The concept of social learning was firstly used by [Shoham and Tennenholtz, 1994], although referring to it as *co-learning*[4]: *"the process in which several agents simultaneously try to adapt to one another's behaviour so as to produce desirable global system properties."*. One of the main contributions of that work is the presentation of the convention emergence problem, characterized by the following dynamics: from a population of agents, n players are randomly selected, and these agents (without any possible means of communication amongst them) choose an action from the m possible following an specific *strategy update rule*; then, the system returns a payoff (calculated depending on the game played, normally a coordination game with a payoff matrix similar to the one in Figure 2.1) with which agents update their *strategy update rule* and the interaction is finished. The system (ideally) continues this process until all agents agree on the same action, considering the convention to have emerged.

In the specific case of [Shoham and Tennenholtz, 1994], the authors focused on the question of how different update rules and other system characteristics affect the even-

[4]Shoham and Tennenholtz [Shoham and Tennenholtz, 1994] were the first ones using the term *co-learning*, also used by [Kittock, 1993], even though the dates might seem misleading. [Shoham and Tennenholtz, 1994] was submitted for publication in 93, when Kittock was writing this other article [Kittock, 1993], which got published earlier.

tual emergence of desirable global system characteristics. The Highest Cumulative Reward (HCR) Rule was object of study in that paper, in a 2-player and 2-actions scenario, analyzing the results when using Coordination Games and the Prisoner's Dilemma games. The HCR rule determines agents which action to choose by observing which action has provided the maximum accumulated reward in the last interactions. We can observe how this strategy update rule is sensible with respect to the memory size of the agents. Their experimental results showed how different parameters that configure the HCR affect to the overall system performance. At this experimentation phase, authors took a decision that would mark the empirical practices on this area of research: they considered the system as converged when it reached **85%** of agents performing the same action[5]. This decision was motivated to save computational efforts. It would seem intuitive that the system would not reverse its state given such high convergence rate, being therefore a perfectly efficient reason to fix the convergence at such level. However, it still was against the strict definition of conventions, which specifies that a convention has to be shared by the whole population in order to obtain the maximum social benefit, otherwise, a subgroup of agents will not be benefiting from it.

Soon after Shoham and Tennenholtz presented their work, Kittock [Kittock, 1993] noticed an important problem in their model. As we said before, these conventions are achieved through social learning, being the interactions amongst agents a key element for the success of the system. Shoham and Tennenholtz did not explicitly consider this problem and used a random interaction approach. Kittock's contribution [Kittock, 1993] is the first attempt to observe the effect of a social network of interaction in the convention emergence problem. Even though the only topologies tested where regular lattices [6] with different neighborhood sizes, those were enough for him to discover the relationship between the ratio of the network and the speed of convergence. However, and following the tradition started by Shoham and Tennenholtz, the empirical convergence rate was fixed to 90%.

Having grounded the base for the convention emergence problem, different extensions have been performed along the years. Without paying attention to the topology of interaction, in the seminal work of Peyton [Young, 1993], a new strategy update rule is described, *External Majority*, where agents update their strategy with a sample of the strategies played by other agents in previous games. For our understanding, such strategy update rule is intrusive for the agents' privacy. Later, in [Walker A, 1995], this strategy update rule was refined to use only information observed by the updating agent.

On top of that refinement, Walker and Wooldridge introduced the convention emergence problem in a Hunter-Gatherer inspired society (previously presented in [Castelfranchi and Conte, 1995]), allowing agents to choose one rule from a set of 4 possible. Moreover, and because of the scenario, agents were not located in a fixed network of interactions; instead they were mobile and located in a grid. Because of the combination of the mobility of agents and the strategy update rule, the system performs a convergence of 100%, becoming the first authors to obtain such achievement.

With the growth of the Internet, networks with similar characteristics to the Internet

[5] An extension of this work is presented in [Shoham and Tennenholtz, 1997b], where the convergence rate is extended to 95% in a subset of experiments (2 of them), remaining the rest (10) equal at 85%.

[6] A *lattice* is a regular network in which all nodes are sequentially connected to a constant number of nodes. The *ring network* and the *fully connected network* are extreme cases of lattices.

Work	Game	Strategy Update	Instruments	Topology	Convergence
[Shoham and Tennenholtz, 1994]	CG & PD	HCR	Memory Restarts	NT	85%
[Kittock, 1993]	CG & PD	HCR	None	RN	90%
[Kittock, 1994]	CG	HCR	Authority	G & T	100%
[Shoham and Tennenholtz, 1997b]	CG & PD	HCR	Memory Restarts	NT	95%
[Young, 1993]	SD	Selective Majority	None	NT	Theoretical
[Sen and Airiau, 2007]	CG & SD	Q-Learning, FP & WoLF-PHC	None	NT	90%
[Mukherjee et al., 2008]	CG	Q-Learning, FP, WoLF-PHC & HCR	None	G	Avg Payoff 3.5
[Delgado, 2002]	CG	Generalized Simple Majority & HCR	None	SF-AB	90%
[Delgado et al., 2003]	CG	Generalized Simple Majority & HCR	None	SF-AB,SF-W & SF-F	90%
[Pujol et al., 2005]	CG	HCR	*Imitation*	Rand, RN, SW & SF-AB	95%
[Urbano et al., 2009a]	CG	Recruitment based on Force with Reinforcement	*Force*	Rand, RN, SW & SF-AB	90%
[Urbano et al., 2009b]	CG	External Majority	*Memory*	Rand, RN, SW & SF-AB	90%
[Walker A, 1995]	CG	7 implementations of *Majority Rules*	None	G	100%
[Savarimuthu et al., 2007b]	UG	Unspecified	None	DN	100%
[Mungovan et al., 2011]	CG	Weighted Replicators Dynamic	Random Interactions	SW + Rand	100%

Table 2.1: Comparative table of Convention Emergence Works. CG = Coordination Game; PD = Prisoner's Dilemma; SD = Social Dilemma; UG = Ultimatum Game. NT = No Topology; G= Grid; T = trees; Rand = Random; RN = Regular Network; SW = Small-World; SF-AB = Scale Free with Albert Barabasi; SF-W = Scale-Free with Walsh; SF-F= Scale-Free Fitness; DN = Dynamic Networks.

captured the attention of scientists, mainly, small-world and scale-free networks. Theoretical scale-free networks have been proved to represent the same characteristics of real social networks [Newman, 2003, Barabasi and Bonabeau, 2003]. In the work of Delgado [Delgado, 2002] it is explicitly analyzed the relationship between the emergence of social conventions and the effect of topologies. In their empirical evaluation they perform experiments with two different strategy update rules: *generalized simple majority* rule and the *highest cumulative reward* rule. In their scenario, pairs of agents have to play a pure coordination game of 2 actions. As the main goal of their work, agents are located in theoretical social networks, namely, Scale-Free Networks constructed with the Albert-Barabasi algorithm[7]. However, as Kittock (and Shoham and Tenneholtz), Delgado fixes convergence rate at 90%. In this case, they observe that the process of convention emergence is accelerated with the usage of Scale-Free networks, because of the cascading effect produced by inherent structure of the network. We will see later how this acceleration is only at an initial phase, as it might happen that certain topological subconventions emerge, specially in scale-free networks, although these cannot be observed when fixing convergence at 90%. In a latter work [Delgado et al., 2003], Delgado's work is extended including different algorithms for the construction of the Scale Free networks (the Albert-Barabasi already studied in [Delgado, 2002], the Walsh model and the Fitness model) to be used in the convention emergence problem. Even though the new extended variety of algorithms to build Scale-Free networks, convergence dynamics are similar and proportional to the density of the network. In a different work [Pujol et al., 2005], and with the scope of accelerating the process of convention emergence, agents are given with socially-inspired mechanisms. Authors proved empirically that by providing agents with a small probability of *Imitation* (copying the action performed by another agent independently of what the strategy update rule dictates), faster convergence rates are obtained in scale-free networks. However, the convergence rates were still fixed at 95%.

Ten years after the first strategy update rule was presented, Sen and Airiau [Sen and Airiau, 2007] decided to incorporate different machine learning algorithms in the convention emergence problem, avoiding in that way the problems inherent to the selection of an appropriate *strategy selection rule*. In that work they did not incorporate any interaction topology, restricting their empirical results to the efficiency of different Multi-agent Learning Algorithms such as Q-Learning [Watkins and Dayan, 1992], WoLF-PHC [Bowling and Veloso, 2002], and Fictitious Play [Fudenberg and Levine, 1998]. The contribution of Sen and Airiau is important as they combine two interesting lines of research in Multi-agent Systems, such as Convention Emergence and Social Norms with Multi-agent Learning. As an extension of the previous work, in [Mukherjee et al., 2008] agents are located in a grid environment, in order to study the topological effects of the social network of interactions, as others ([Walker A, 1995]) did before. In both works, they consider a convention to have emerged when the average payoff of the population is at 3.5 (out of a maximum of 4, and a minimum of -1).

In a renewal of interest for the Convention Emergence problem, in the last years the MAS community has focused on discovering different mechanisms that could ac-

[7] Albert-Barabasi algorithm tuned with the following parameters $m_0 = 4$, $m = 2$, $p = q = 0.4$.

celerate the emergence, specially in complex networks. In the work of Savarimuthu et al. [Savarimuthu et al., 2007b], authors describe the emergence of norms in an scenario where the network of interactions dynamically change as the agents move in a grid world, being their neighbours in the network those agents geographically closer to them in the grid at an specific moment in time. Eventhough this work might remind us to others [Walker A, 1995], authors here focus on the dynamic aspect of the network of interactions and how this affects to the process of emergence, which in this case is fixed at 100%.

Another recent work [Mungovan et al., 2011] propose an interesting scenario where agents have a fixed interaction network (small-world) and authors provide agents with the possibility of having random interactions. Their solution to the random interactions is very elegant as the random interactions are not chosen uniformly from the population, but, with a weighted measure considering the distance in the network (i.e. agents that are closer in the network are more likely to interact). Their results, together with those presented in [Walker A, 1995, Pujol et al., 2005, Savarimuthu et al., 2007b], suggest that a random interactions produce faster convergence results than when the interaction topology is fixed. However, the type of scenario we are interested in (virtual societies constructed on top of a real social network) cannot assume the possibility of intense dynamic changes on the underlying network.

Trying to provide agents with other mechanisms, Urbano et al. [Urbano et al., 2009a] investigated how a new strategy update rule (*Recruitment based on Force with Reinforcement (RFR)*), which promotes a dynamic creation of a hierarchy speeds up the convention emergence. Their proposed strategy update rule functions by providing agents with a measure of *force* that increases with every successful interaction, and in case of unsuccessful interactions, the agent with the lower force copies the strategy and force of the winner. We find this function similar to the SLACER algorithm [Hales and Arteconi, 2006]. Authors prove the efficiency of their new strategy update rule on different topologies: regular graphs, a unconventional implementation of a random graph with uniform degree distribution, scale-free networks (developed following the Albert-Barabasi model), and Small World (constructed using the Watts-Strogatz model). However, convergence rate is still fixed to 90%. Their results are in concordance with those obtained by Kittock in [Kittock, 1994] where he proved the efficiency of trees structures, that are those dynamically created with the concept of force.

Providing a slightly different framework than the classical convention emergence, Salazar et al. [Salazar et al., 2010] propose a variation of the coordination game, implementing a Language Coordination Game similar to [Lakkaraju and Gasser, 2008]. In their game, agents have to match words with concepts, creating a huge convention space. Their solution is founded on a *spreading protocol* where agents can communicate part of a local solution together with a truthful valuation of the solution. Therefore, agents aggregate into their solutions the bests pieces of solutions received and then communicate them again with a truthful measure. Because of the communication issue, their framework is fragile against the presence of untruthful agents with different preferences than their collaborative agents. Authors demonstrate the robustness of their *instrument* in different topologies (scale-free and small world), reaching convergence

rates of 100% in most of the cases. Even though the success of their approach, they assume the existence of a fully coperative community, which in our case does not have to be the case. Conventions have to be reached by themselves, as it is in the common interest to be reached.

The general conclusion that we can extract from the revised work on the literature is that only those authors that provided agents with instruments (generally instruments that access private information of agents) achieve complete convergence in complex networks. The rest of work which do not incorporate any extra mechanism state the convergence results with a rate of 90%, that as we have claimed previously, it cannot be considered as a convention to be emerged. A summary of the review works in the literature together with some of the characteristics analyzed can be seen in Table 2.1.

In the following chapters we will overcome this prefixed limit, observing the dynamics of the system after that point.

2.3 Essential Norms and the Emergence of Cooperation in Artificial Social Systems.

As we mentioned previously in this chapter, there are norms that solve different problems (represented with different games) than conventions. These games are represented by the Prisoner's Dilemma or the Public Good Game (depending on the amount of players). In general, in these type of games, the individual's self-interest is in conflict with the social interest, and norms establish a balance, motivating agents to play the prosocial behaviour. These type of games generally represent what in the literature is known as the *Emergence of Cooperation* [Axelrod, 1981]. This problem has been treated from several areas of research such as philosophy [Bicchieri, 2006], economy [Fehr and Gachter, 2002, Dreber et al., 2008], politics [Axelrod, 1986], and computer science[Kollock and Smith, 1996, Shoham and Tennenholtz, 1997a].

"The Tragedy of the Commons", firstly described in [Hardin, 1968], describes the situation where potentially everyone can highly benefit from a common resource. However, the temptation is to selfishly enjoy the common resource without contributing to it.

In order to achieve and maintain cooperation, researchers have looked for socially-inspired techniques that can be implemented into the agent architectures and obtain a macroscopic result. Along this section we will review the most influential techniques that can be suitable for decentralized virtual societies, focusing on the achievement of cooperation.

2.3.1 Socially Inspired Mechanisms for Cooperation Emergence

Tags

One of the most studied mechanisms is *Tags*. The term was introduced by Holland in [Holland, 1993], and it is ideally conceptualized to allow agents to have a certain external visible marking recognizable by the rest of the population and relating the tag

with a specific behaviour. A good example of a *Tag mechanism* is to be found at the military stripes, allowing any member of the military to recognize the rank (and associate it with certain protocols of interaction) towards another by observing its stripes.

By allowing agents to carry tags, they recognize features that can associate towards others' cooperative or defective strategy. It has been showed that incorporating a tag-based donation system (donating to those agents with similar tags) can sustain the emergence of cooperative behavior [Riolo et al., 2001]. Nevertheless, the model is based on a strong assumption: agents change their tags by accessing (and copying) other more successful agents' tag. We consider such information (others' strategy and its success) to be private, and not accessible to other agents, at least in the type of scenarios we envision, where each agent might belong to a different owner, and therefore, this type of information access is unrealistic.

In addition, in these models tags evolve by allowing agents to substitute their neighborhood for the neighborhood of the most successful agents, producing a dynamic variation of the network as well as an unrealistic access to private information of agents (such as their network of interactions).

Other models [Hales and Edmonds, 2005, Hales and Arteconi, 2006] have tried to apply a tags based mechanism in P2P networks, obtaining higher cooperation rates than when tags where not used. Other models have been inspired by Hales' SLAC algorithm (copying the neighbours and strategy of the most successful known agent) like Araujo's [Araujo and Lamb, 2008] *Memetic Networks* (inspired in the concept of *memes*: information that is propagated through copies in a cultural setting), or Griffiths' et al. [Griffiths and Luck, 2010] and the integration with *Context Awareness*.

Even though some variants of the tags mechanism have been analyzed [Matlock and Sen, 2009], we find unrealistic to allow agents to access others' private information, which necessary for tags to function (success of the opponent agent, and strategy of interaction) and copy it for their own success. Moreover, in tag based mechanism, agents reproduce and die as genes do in a genetic algorithm. We are more interested in finding online mechanisms that rely on the present and public actions of agents without having to sacrifice or reproduce.

Partner Selection

The private information access problem that concerns us the most with respect to *Tags* mechanisms is avoided in another studied mechanism: *Partner Selection*. The main difference with respect to tags is that this mechanism allows agents to choose whom they want to interact with, thus reducing the risks associated to cooperation. *Tags* allow agents to identify other agents as similar (comparing their tags) and then interact accordingly; however, with *partner selection* agents need to explore and then start the exploitation by interacting repeatedly with those agents that have obtained the most satisfactory results. This type of mechanism generates the evolution of a social network, where an agent interacts more often with those that it considers more suitable. Depending on the agents' tolerance towards defection and the costs of refusal and social isolation, the resulting social network of interaction might be very detached, resulting in higher rates of endogamy [Ashlock et al., 1996].

From the point of view of defectors, one of the problems is that in large populations

they might avoid punishment by roving from group to group, while they might result ostracized in smaller populations. Some simulation models of real phenomena have proven [de Vos et al., 2001, Tosto et al., 2006, Nosratinia et al., 2007] that by allowing partner selection, higher group payoffs are obtained, increasing the rate of cooperation while avoiding free-riders.

However, these models implementing partner selection also implement the possibility of interaction rejection between agents, producing a segregation on the population.

Reputation

One mechanism that has been thoroughly studied is *Reputation*. This mechanism emerges when agents can communicate amongst them and can express and reason about other agents previous experiences. MAS researchers have produced a number of different models to reason about trust and reputation issues [Sabater and Sierra, 2005], in a centralized or decentralized way. The last models (decentralized) are those that we are more interested in. Reputation mechanisms [Ohtsuki and Iwasa, 2004, Nowak and Sigmund, 2005] help in establishing indirect reciprocity ("I help you if you help somebody else"). However, these type of systems are fragile against the existence of unreliable agents.

Reputation can be used to promote partner selection (by spreading good evaluations about an agent and therefore everyone will want to interact with it) or to punish (by spreading bad evaluations about such agent producing ostracism).

Some MAS scientists have proven that by introducing reputation mechanisms in P2P-like scenarios, cooperation is sustained [Hales, 2002, Banerjee et al., 2005], ensuring a good performance of the system.

Authority

The existence of a centralized authority that regiment the behaviour of agents is an efficient mechanism to obtain norm-compliance [Cardoso and Oliveira, 2008, Cardoso and Oliveira, 2009, Garcia-Camino A, 2006, Grossi et al., 2007]. The success of the authority is found on the recognition from agents of an entitled figure that objectively recognizes the difference between what is right or wrong, and (ideally) punishes what is not considered as a pro-social behaviour.

Our main concern against the *Authority* is that it requires role-specialization, where some agents have special abilities over the others. The type of scenario that we deal with is self-regulated, which assumes that all agents have the same potential possibilities to perform actions over the other agents. The MAS scenarios where *authority* has been proven are small and controllable environments, however, we can think of other types of scenarios (like P2P networks) where controlling all interactions amongst agents is totally unfeasible.

Authority has also been implemented [Aldewereld et al., 2005] into the agents mind: instead of having an entitled agent controlling the correct interactions, agents are initialized with normative protocols of interactions, called *landmarks*, which restrict the agents freedom, establishing the actions to be taken in order to achieve certain

tasks in a normative manner. The main drawback associated to landmarks is the difficulty for agents to adapt to environmental changes, which has to be done by changing agents' internal landmarks by their programmers.

In general, we can observe how the success of most of the previously explained mechanisms is based on the potential repercusions on the agents payoffs, therefore, the punishment associated. When using *tags* or *partner selection*, agents fear the potential ostracism [de Pinninck et al., 2007a]. *Reputation* affects indirectly the utility of an agent, by restricting the number or decreasing the quality of interactions(due to the bad reputation). And, with *authority* is implemented, agents fear the fines imposed by authority when norm violations are observed. Because of all these reasons, we focus on punishment as one important mechanism that ensures cooperation.

2.3.2 Punishment

Allowing agents to react against the behaviour of wrongdoers is the scope of *Punishment*. On the other hand, by receiving punishments, agents' utility can be affected by the others punishments. Nobel prize Elinor Ostrom [Ostrom, 2000] gives an evolutionary perspective to the emergence and maintenance of the cooperation norm in collective action dilemmas, discussing the relationship of the maintenance of norms with the existence of monitoring and sanctioning mechanisms that reduce the probability of free-riding. Another Nobel prize Gary Becker also related the existence of a punishment mechanisms with the maintenance of cooperation [Becker, 1968]. Nonetheless, Becker's view describes a *homo-economicus* type of agent, which is only motivated by utilitarian aspects. The pessimistic conclusion of many researchers [Hardin, 1968] is that coercion by a strong external authority is necessary in order to ensure cooperation.

However, Ostrom's theory suggests the existence of other type of mechanisms that sustain cooperation other than the utilitarian damage that comes with punishment. In [Ostrom, 2000], she discusses the laboratory empirical evidences that are hard (or impossible) to explain following Becker's homo-economicus theory, or zero-contribution thesis.

Ostrom defends that for the evolution and maintenance of norms, other types of agents different to the rational egoist agent have to be present in the society, and she defines two of them: *"conditional cooperators"* and *"willing punishers"*. The former are willing to cooperate to the common good (even though the rational strategy would move the agent to reduce its contribution to the minimum, even to zero). However, this type of agents is also affected by the behaviour of others, which might provoke a collapse in the cooperation strategy followed by the conditional cooperator. On the other hand, the willing punishers are agents that are *"willing, if given the opportunity, to punish presumed free riders through verbal rebukes or to use costly material payoffs when available"*. With the existence of these types of agents, a more robust compliant and norm-defending society can be obtained.

Several works in Multi-agent Systems have considered the existence of punishment mechanisms to ensure norm-compliance [de Pinninck et al., 2007b, Blanc et al., 2005, Grossi et al., 2007, Bazzan et al., 2008]. In [Pasquier et al., 2006] a classification of the possible types of sanctions that can be present in a MAS, together with the styles of sanctions and the scope of sanctions, is presented.

By allowing peer punishment, on contrast to that achieved through a centralized institution, agents have the option to violate norms in those situations that they find it necessary. This intelligent norm-violation allows society as a whole to achieve self-organization.

Normative BDI Architectures

Some authors [Castelfranchi et al., 2000] already noticed about the necessity of intelligent norm violation. They describe a theoretical architecture of agent that have a number of characteristics that allow an agent:

- to know that a norm exists in the society and that it is not simply a diffuse habit, or a personal request, command or expectation of one or more agents;
- to adopt this norm impinging on its own decisions and behaviour, and then
- to deliberately follow that norm in the agents behaviour, but also
- to deliberately violate a norm in case of conflicts with other norms or, for example, with more important personal goals; of course, such an agent can also accidentally violate a norm (either because it ignores or does not recognize it, or because its behaviour does not correspond to its intentions).

Even though the accurate theoretical policy described in [Castelfranchi et al., 2000], no agent implementation was done. On the other hand, a more complete description of an agent architecture based on a BDI architecture is presented in [Dignum et al., 2000], which integrate norms into the agents decision making, specially on the process of generating (candidate) intentions. With an such type of architectures, agents' behaviour is different (and norm regulated) than that of utility-maximizing agents.

[Boella and Lesmo, 2002] sketched a BDI agent architecture that cope with game-theoretical concepts of norms, mainly, punishment. A key element for their theoretical model to function is the existence of a *"normative agent"*, which imposes obligations to other agents, observe the fulfillment of norms, and apply sanctions to norm-violators, similar to the *"willing punishers"* described in [Ostrom, 2000].

As we will explain along this thesis, there is a clear necessity of a modular BDI agent architecture that can deal with normative related actions such as punishment and integrate them correctly into the agents decision making. On the other hand, this agent architecture has to be designed in such way that agents can intelligently apply punishments to other norm-violating agents. With a correct implementation of a BDI normative architecture which reasons about norm compliance but also about metanorm compliance (apply punishments), intelligent norm violation can be achieved.

Other researchers have investigated a negotiation approach [Boella et al., 2009] to the emergence of norms. In this work, agents negotiate about all the aspects related to the norms, like the social goal to be achieved with the fulfilment of the norm, the obligations and sanctions. Moreover, this model is theoretically designed for online game scenarios, like The Sims Online or Second Life. And alternative to the BDI models to achieve norm consensus is that presented in [Joseph and Prakken, 2009].

Instead of centering on the emergence, other proposals [Tinnemeier et al., 2010] have been envisioned to facilitate the runtime modification of norms. Such framework assumes a whole cooperative population, that will accept any changes in the normative code produced by any other agent. Similar to the work of Tinnemeier, in [Meneguzzi and Luck, 2009] authors present an agent architecture able to modify plans at runtime in reaction to newly accepted norms. Their main contribution is the introduction of plan manipulation strategies to enable reasoning about norms and to ensure compliance with newly accepted norms. Authors focus on the intelligent compliance of norms rather than in the emergence of these. They present an elegant framework where norms are obligations and prohibitions over certain states of the world or certain actions, which are specified with a validity period (if the norm says to pay your bills, it is impossible to identify a norm violation if there is no deadline to abide by the norm). Their framework can generate new plans to enable agents to comply with norms, and remove the plans when the norms are no longer relevant.

Another important contribution in the literature of normative agent architecture is EMIL-A [Andrighetto et al., 2010a]. EMIL-A is an agent architecture born with the scope of representing the process of norm innovation, and how this process is affected by different societal levels, and the relations amongst them. Specially, they focus on the recognition of social norms in open societies where there is no central authority. The architecture is built using a BDI approach, where the micro-interactions amongst agents derive to a macro-behaviour (norms). Their theoretical architecture is also proven with a simulation framework, where agents recognize the existence of norms and behave according to that recognition. However, agents were still not given with the tools and mechanisms to reason about the dynamic state of norms, and how they might disappear (and loose scope) along time.

In this thesis we will present an agent architecture (EMIL-I-A), constructed on top of the BDI EMIL-A architecture, that is able to interpret normative cues (such as compliance, violations, sanctions and punishments) into the agents decision making. Moreover, this architecture integrates an adaptive vision of agent which can adapt to environmental and normative changes and obtain intelligent norm violation. Finally, this architecture is given with an extra mechanism that allows agents to dissociate compliance with punishment, called *Internalization*. Our EMIL-I-A Architecture fits with the operational approach described in [Lotzmann and Möhring, 2009].

Chapter 3

Conventions: Trespassing the 90% Convergence Rate

Conventions are a special type of norms, used as a mechanism for sustaining social order, increasing the predictability of behavior in the society and specify the details of those unwritten laws. Following Coleman's theory[Coleman, 1998], conventions emerge to solve coordination problems, where there exist no conflict between the individual and the collective interests, as what is desired is that everyone behaves in the same way, without any major difference on which action agents are coordinated. Therefore, the selection of the focal action in such norms is arbitrary. One clear example of these kind of norms is the selection of which side of the road to drive on, both options are equally good as long as all drivers agree. Examples of conventions pertinent to MAS would be the selection of a coordination protocol, communication language, or (in a multitask scenario) the selection of the problem to be solved.

Research from several communities (multi-agent systems, physics, or economy) have studied a restricted configuration of parameters, giving special importance to the underlying interaction topology [Kittock, 1993, Shoham and Tennenholtz, 1997a, Urbano et al., 2009a, Delgado, 2002, Delgado et al., 2003], and leaving all the other parameter configurations unexplored.

Moreover, and in order to save computation efforts, most of these research considered a convention as emerged when 90% of the population had converged to the same convention. However, 90% convergence cannot, by definition, be considered a convention, as a convention needs to be shared by the complete population. Therefore, we have decided to extend the 90% convergence rate to 100% in order to be consistent with the definition.

In this work we will provide multi-dimensional analysis of several factors that we believe determine the process of convention emergence, such as:

- the population size.
- the underlying interaction topology.
- the strategy selection rules.

- the learning algorithm.
- the agents memory size.
- the amount of players per interaction.

The contribution of this chapter is centered on the achievement of full convergence, and a detailed analysis of the search space of parameters.

3.1 Model

The social learning situation for norm emergence that we are interested in is that of learning to reach a social convention. We adopt the following definition of a social convention from the definition of a social law [Shoham and Tennenholtz, 1997a]: *A social law is a restriction on the set of actions available to agents*. A social law that restricts agents' behavior to one particular action is called a social convention.

For the sake of generalization, our framework is built with the most accepted convention emergence model (used by [Kittock, 1993, Walker and Wooldridge, 1995, Shoham and Tennenholtz, 1997a, Delgado et al., 2003, Mukherjee et al., 2007, Sen and Airiau, 2007, Villatoro et al., 2009]). We represent the interaction between two agents as an n-person m-action game.

for *a fixed number of epoch* **do**
  $A \leftarrow N$ `// Initialize the available agents with all the population`
  repeat
    select randomly an agent from the available agents: $i \in A$;
    select randomly a neighbor of i that is available: $j \in A \cup \{j \,|\, (i,j) \in \mathscr{G}\}$;
    remove i and j from the set of available agents: $A = A \setminus \{i,j\}$;
    ask i to select an action r in $\mathscr{A}_r$;
    ask j to select an action c in $\mathscr{A}_c$;
    send the joint action (r,c) to both i and j for policy update;
  until *no pair of agents is available:* $\nexists (i,j) \in A^2 \,|\, i,j) \in \mathscr{G}$;
end

Algorithm 1: Interaction protocol.

The model runs as specified in Alg. 1: at each time step, each agent is paired with another agent from the population and chooses from one of several alternatives (the choice can either be that of an action to execute, which we use in this paper, or a particular state to be in). In our case a social convention will be reached if **all** the n agents are in the same state or choose the same action, i.e., the actual state or action chosen as a convention is not important.

Based on the underlying assumptions of the convention emergence problem, all actions (or states) are potentially of the same value, as long as all agents converge to the

same action (i.e. given the set of actions=$\{A, B\}$, converging to action A produces the same utility to agents than converging to action B).

We model agent environments by networks, where each agent is represented by a node and the links in the network represent the possibility of interaction between nodes (or agents).

We consider the following two different agent network topologies or environment types:

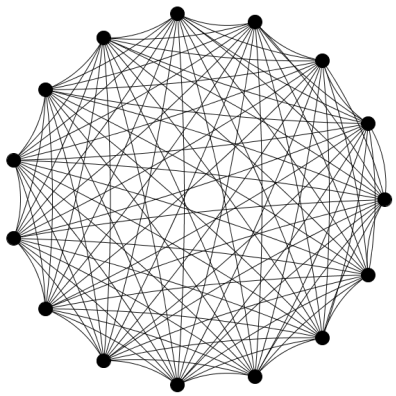
1. a *one-dimensional lattice* that provides a structure in which agents are connected with their n nearest neighbors. Different values of the neighborhood size (n) produce different network structures. For example, when $n = 2$ the network will have a ring structure (as in Figure 3.1(b)) and agents will only be connected with their immediate neighbors on either side, corresponding to a ring topology. On the other hand, when $n = PopulationSize$, the connections result in a fully connected network (as in Figure 3.1(a)) where each agent is connected with all other agents. The motivation for the usage of this type of networks is to simulate the ordered structure that many computational systems (memory clusters, sensor networks, etc..) are organized.

2. a *scale-free network*, whose node degree distribution asymptotically follows a power law, meaning that there are many vertices with small degrees and only few vertices with large degrees. This makes the network diameter (average minimum distance between pairs of nodes) significantly smaller than *one-dimensional lattices* with small neighborhood sizes. Scale-free networks do represent the topology of social networks [Albert and Barabasi, 2002], and we find valuable to have results at least in synthetic scale-free networks. An example of this topology can be seen in Fig. 3.1(c). The scale-free networks used in this work are generated with the BA algorithm [Barabási and Albert, 1999], with $m_o = 2$, $m = 1$, $p = 1$ and $q = 0$, as in other works such as [Delgado et al., 2003, Salazar et al., 2010].

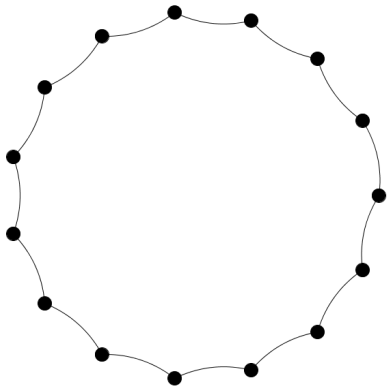
3. to capture some typical real-world scenarios, e.g., a community of closely knit researchers and their students, we use a rather novel network topology, namely the *fully connected stars network*: such a network has a relatively small number of hubs or core nodes which are fully connected forming a clique, and each of these core nodes is also connected with a number of leaf nodes (an example can be seen in Figure 3.1(d)). This topology possesses some interesting characteristics to understand the topological dynamics of convention emergence.

Agents cannot observe the other agent's current decision, or immediate reward, and hence cannot calculate the payoff for any action before actually interacting with the opponent.

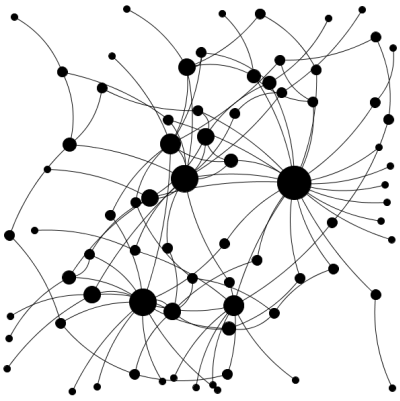
Agents use a learning algorithm to estimate the worth of each action. Agents will choose their action in each interaction in a semi-deterministic fashion. A certain percentage of the decisions will be chosen randomly, representing the exploration of the agent, and for the rest of the decisions, the agents deterministically choose the action estimated to be of higher utility. In all the experiments presented in this work, the exploration rate has been fixed at 25%, i.e., one-fourth of the actions are chosen randomly.



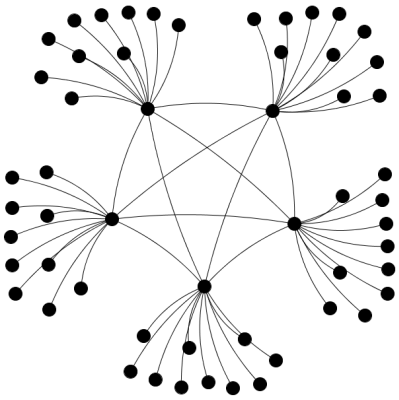
(a) Fully Connected Network



(b) Ring Network or One-dimensional Lattice with Neighborhood Size 2



(c) Scale-Free Network



(d) Fully Connected Stars Network

Figure 3.1: Underlying Topologies

The learning algorithm used here is a simplified version of the Q-Learning algorithm [Watkins and Dayan, 1992]. The Q-update function for estimating the utility of an action is:

$$Q^t(a) \leftarrow (1-\alpha) \times Q^{t-1}(a) + \alpha \times reward, \tag{3.1}$$

where *reward* is the payoff received from the current interaction and $Q^t(a)$ is the utility estimate of action a after selecting it t times. When agents decide not to explore, they will choose the action with the highest Q value. The reward used in the learning process is a proportional reward obtained from the system after each interaction.

We have used two different learning modalities: (a) in the *Multi learning* approach both interacting agents use the payoff to update their memory and action estimate, (b) in the *Mono learning* approach, however, only the first agent selected, and not the second one, updates its memory and action estimate after an interaction. Each agent interacts exactly once per time step in mono-learning, whereas in the multi-learning mode, different agents interact different times in the same time step because of random partner selection.

3.2 Overpassing the 90%

The first experiments performed have the scope of proving how convergence evolves in the convention emergence process, as done in [Delgado et al., 2003, Kittock, 1993, Mukherjee et al., 2007, Sen and Airiau, 2007, Shoham and Tennenholtz, 1997a, Walker and Wooldridge, 1995]. However, this time we have fixed the rate of convergence to 100%.

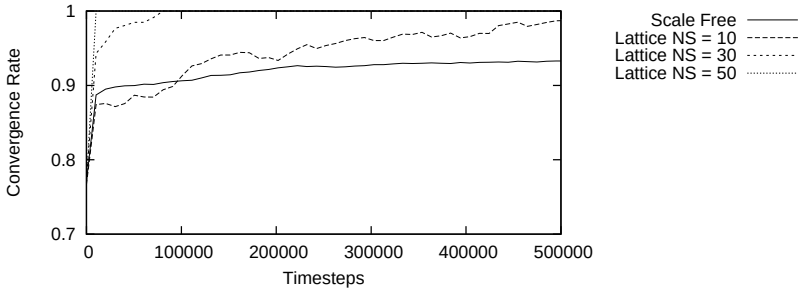
Our experimental results confirm those obtained by others with the same parameters, although, as we venture further than 90%, we discover different dynamics in those ranges.

It can be seen in Figure 3.2 that convergence rates grows really fast at the beginning of the execution as agents do not have any clear preference and they will converge through the learning obtained by interacting. In certain topologies, the convergence rates suffer from a delay which might be produced by the existence of subconventions.

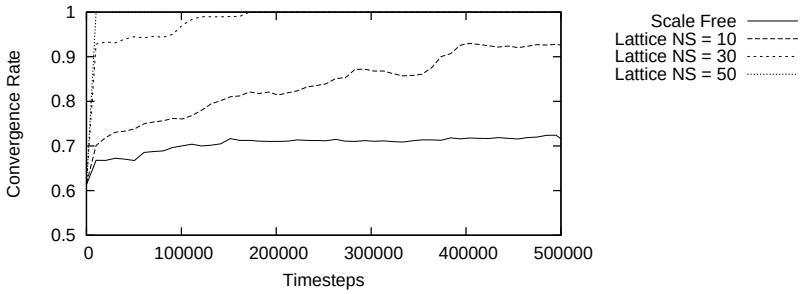
We hypothesize that subconventions are facilitated by the topological configuration of the environment (isolated areas of the graph which promote endogamy) or by the agent reward function (concordance with previous history, promoting cultural maintenance). While the former can be studied by analyzing the dynamics of convention emergence of the different presented topologies, in order to test the latter a new reward function that captures the agents' previous history has to be designed.

3.3 Is cultural maintenance the source of subconventions?

We hypothesize that history of interaction is instrumental for norm evolution. Learning algorithms incorporate the history of interactions into their decisions, but reward metrics are typically static and independent of the agent histories. Norm evolution is



(a) 2 actions



(b) 5 actions

Figure 3.2: Convergence rates with different topologies and actions set

dependent upon the exertion of social pressure by the group on aberrant individuals. It is through learning via repeated interactions that social pressure is applied to individuals in the group. However, a reward metric based on the current interaction does not necessarily model the full context or capture the persistent nature of social pressure in human societies.

In particular, society often uses past history to judge individuals and hence actions have future consequences in addition to immediate effects. Accordingly, we propose a reward structure based upon the agent's interaction history as a more appropriate alternative to the single interaction reward metric normally used in the literature. In our model agents are rewarded based upon the conformity of action between two agents, such that the agent who has the larger of the majority action choice in the stored interaction history of the agents receives higher reward. Hence, both interaction agents' history of actions are used to calculate each individuals' payoff from an interaction. We investigate how this history, and in particular, its size (memory size) affects different types of social structure.

We have designed a memory based reward function that captures the past actions of the interacting agents into the calculation.

```
   /* First, we select the majority action                    */
 1 TotalAActions = A_1 + A_2;
 2 TotalBActions = B_1 + B_2;
 3 if TotalAActions < TotalBActions then
 4 |   MajorityAction = B;
 5 if TotalBActions < TotalAActions then
 6 |   MajorityAction = A;
 7 if TotalAActions == TotalBActions then
 8 |   MajorityAction = RandomlyselectedbetweenAorB;
 9 /* Then, we calculate the reward depending on the agents
      action selection and on the majority action             */
10 if Action_1 == MajorityAction then
11 |   reward_1 = MajorityActions_1 / TotalMajorityActions
12 else
13 |   reward_1 = 0
14 end
```

Algorithm 2: Memory Based Reward Function.

When two agents interact, the instantaneous reward that an agent receives is calculated based on the action it selected and the action history of both agents as shown in Algorithm 2 (the algorithm calculates reward for agent 1 and assumes only two actions available per agent, but can be easily extended to an arbitrary number of actions). Here A_x and B_x are the number of A and B actions in memory that agent x has taken, $Action_x$ is the last action taken by agent x, and for which it is rewarded, $MajorityAction$ is selected to be whichever action is selected more frequently by the two players combined, $MajorityActions_x$ is the number of actions equal to the majority action that agent x has previously taken, and $TotalMajorityActions$ is the number of times the majority action

was chosen by both players in their finite histories.

3.3.1 Experiments

To evaluate the rate and success of norm emergence we ran experiments with different societal configurations by varying the following system and agent properties:

Memory Size: To be able to analyze the effects of memory sizes, we vary the number of past interactions stored by an agent,

Population Size: We study the effects of scale-up by varying population size.

Neighborhood Size: We study how different neighborhood sizes in a one dimensional lattice affect the process of emergence of conventions.

Underlying Topology: We observe the dynamics of the process of emergence of conventions depending on the underlying network topology.

Learning Modalities: We compare how conventions are reached with different learning modalities, namely, one or both agents learning from an interaction.

Number of Players: We analyze the effect of different number of players participating in each interaction on the speed of convention emergence.

Actions Set: We study the effect of the search space, namely the number of options available to agents, on the convention emergence process.

Results reported here have been averaged over 25 runs. Agents are initialized with uniformly random memories, and initially are unbiased in their action choice. We conclude that a social convention has been reached when 100% of the population choose the same action.

3.3.2 Effect of Memory Size

In this set of experiments, we want to observe the effect of different memory sizes on convention emergence. We fix the population size at 100 agents. We present the convergence times for different memory sizes in Figure 3.3.

First experiments are performed on fully connected networks, and the results are shown in Figure 3.3. For both learning modalities, convergence times increase proportionally with memory size. The relationship between memory size and convergence time is explained by observing the configuration of the reward function and the learning algorithm. Each action in memory gets a relatively high reward for smaller compared to larger memory sizes (refer to the reward function defined in Algorithm 2). The learning algorithm, therefore, receives larger reinforcements for the actions performed for smaller memory sizes, resulting in faster convergence. Convergence is accelerated in this situation because higher rewards have a larger impact on the Q value updated by the learning algorithm represented in Formula 3.1. On the other hand, when dealing with higher memory windows, the proportional reward is much smaller, and therefore, the reinforcement will be smaller. Due to this smaller reinforcement, a higher number

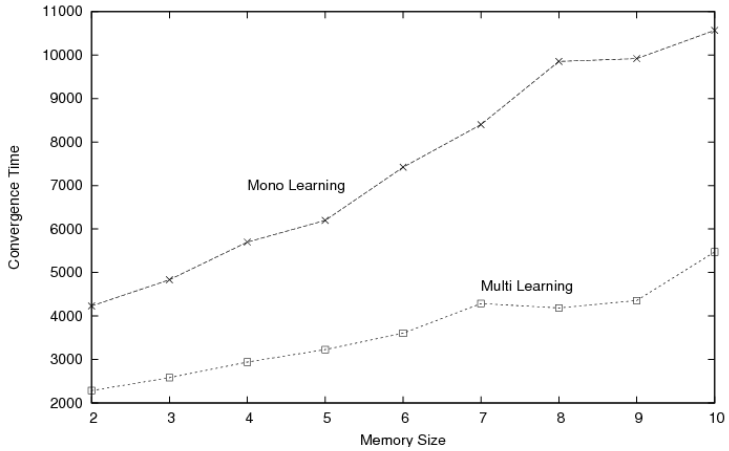


Figure 3.3: Effect of Memory Size in Convergence Time on a Fully Connected network. (100 Agents).

of interactions, and hence higher number of timesteps, will be needed to reinforce that action to same degree, thereby increasing convergence time.

Moreover, and observing the learning modalities, the mono-learning approach takes longer to converge than the multi-learning approach. A part of this difference is explained by the fact that the average number of learning interactions in a multi-learning approach is twice that of the mono-learning approach for the same number of time steps. There is, however, an additional clear trend of accelerated learning when both agents are learning from the same interaction.

We extend the experiment to the other presented network topologies. The results show that larger memory sizes consistently increase convergence times for all the topologies, showing a stronger impact in the scale-free and fully connected stars networks than in the one-dimensional lattice.

In Figure 3.4 (note that the y-axis is in a logarithmic scale), we can observe the relative performance of different topologies for different memory sizes with the mono-learning approach. During this experiment, we limited the execution of the simulations to one million timesteps. We observe that the Fully Connected Stars network takes the most time to converge, followed by the Scale-Free network. For both the Scale Free and the Fully Connected Networks we can observe that the convergence time increases with increasing memory size. These inefficiencies are largely due to more time taken to break or resolve conflicting subconventions that form with scale-free and fully connected stars networks but not for fully connected networks (see Section 3.3.4). Accordingly, the fully connected network scales up much better with increasing memory size.

These first experiments provide us with the hint that topologies have an strong impact on the delay of full convergence, specially in Scale-Free and Fully Connected Stars networks.

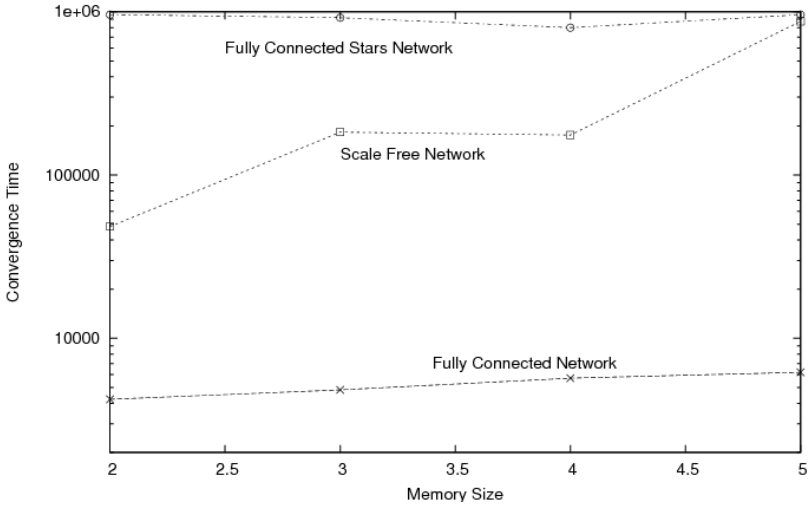


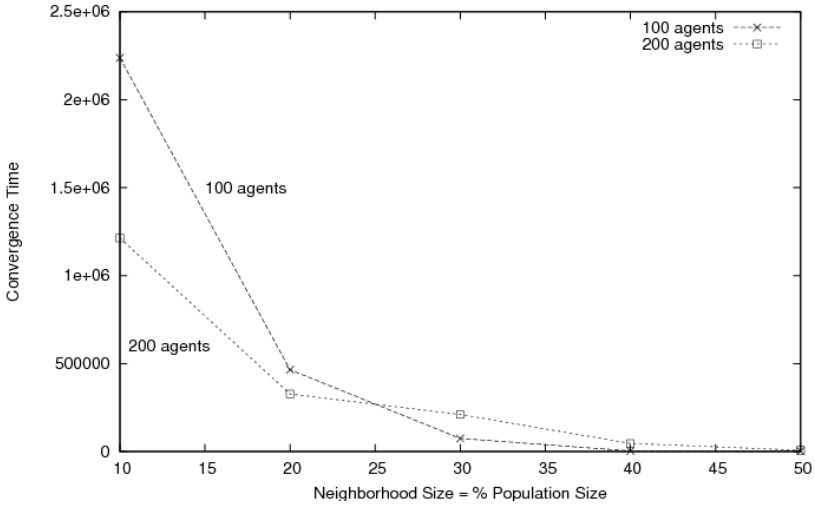
Figure 3.4: Topologies Comparison with different Memory Sizes with Mono Learning Approach.

3.3.3 Effect of Neighborhood Size

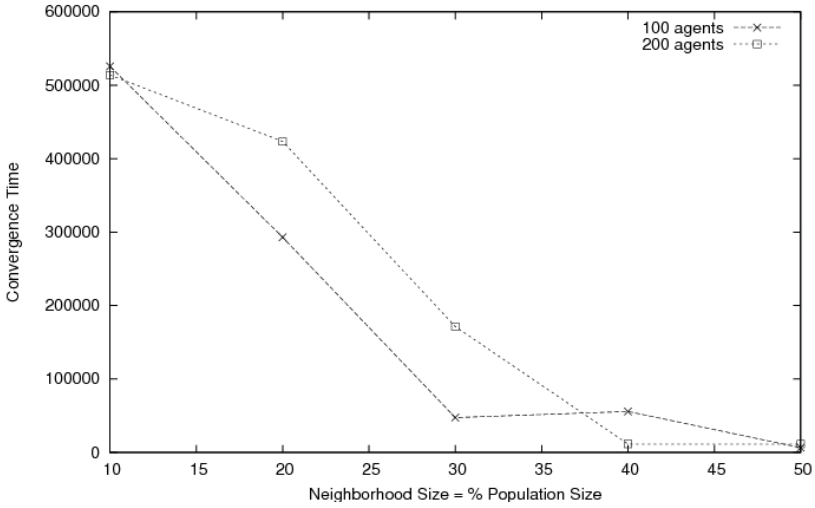
To observe the effect of neighborhood size, we use a one-dimensional lattice (as scale-free networks and fully connected stars predetermine the neighbors for each node). In these experiments, we fixed a memory size of 5. Figure 3.5(a) shows a comparison of convergence times for different neighborhood sizes, measured as percentages of the population size, in a multi learning approach.

We can see that when increasing the neighborhood size, the convergence time is steadily reduced until it stabilizes after a certain neighborhood size. This effect is due to the topology of the network. When the one dimensional lattice has a small neighborhood size, on average, the diameter of the graph[1] is high and therefore agents have a relatively higher amount of local interactions. These local interactions might promote different conventions in different parts of the network due to the inherent structure of the convention emergence problem: one convention from the possible ones has to be adopted by interacting with your neighbours, and as the topology restricts agents to more immediate neighbours (promoting endogamy), they can develop a different convention than agents in another part of the network. These subconventions need to be broken within a certain number of timesteps to achieve the global convention. In the case of larger neighbourhood sizes, these endogamy has a softer effect, and therefore the topology does not promote such strong metastable subconventions. It is also interesting to note that for smaller neighborhoods, larger populations exhibits much faster convergence.

[1]The diameter of a graph is the largest number of vertices which must be traversed in order to travel from one vertex to another



(a) Multi Learning



(b) Mono Learning

Figure 3.5: Convergence rates with different Neighborhood Sizes in a One Dimensional Lattice

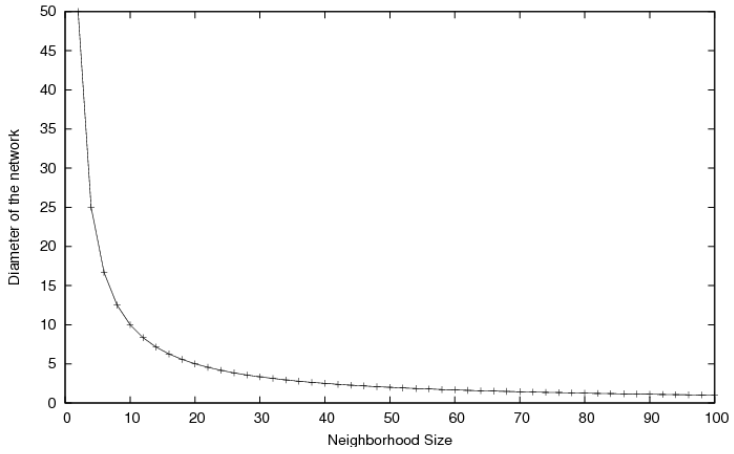


Figure 3.6: Diameter relation with Neighborhood size in a One Dimensional Lattice with population = 100

Similar convergence results are also obtained with the mono learning approach shown in Fig. 3.5(b). When the neighborhood size crosses about 30% of the population size, the convergence time does not significantly decrease anymore. The relation of the neighborhood size and the diameter follows a geometric distribution and is shown in Figure 3.6. We see that when neighborhood sizes cross 30% of the population size the diameter of the network is no longer significantly reduced, and hence the convergence times are also not significantly reduced any further.

This experiment convinced us of the importance of the topology, but also lead us to think about the differences between the learning modalities and their different effect.

3.3.4 Effect of Learning Modalities

In this set of experiments, we want to observe the difference in convergence times with the two learning modalities for different topologies. We first compare results of the two learning modalities in a one-dimensional lattice with 100 agents (see Figure 3.7, where the y-axis is drawn on a logarithmic scale). For smaller neighborhood sizes, i.e., when the network diameter is high, multi-learning takes longer to converge than mono-learning. After reaching the point where the diameter is no longer affected by the neighborhood size (as discussed before, this happens when the neighborhood size is about 30% of the population size), the multi-learning performs better. The reason for this interesting phenomenon is the creation of local subconventions with multi-learning when the neighborhood size is small.

When agents have a small neighborhood size, they will interact often with their neighbors, resulting in diverse subconventions forming at different regions of the network. With the multi-learning approach, agents reinforce each other in each interaction. Such divergent subconventions conflict in overlapping regions. To resolve these

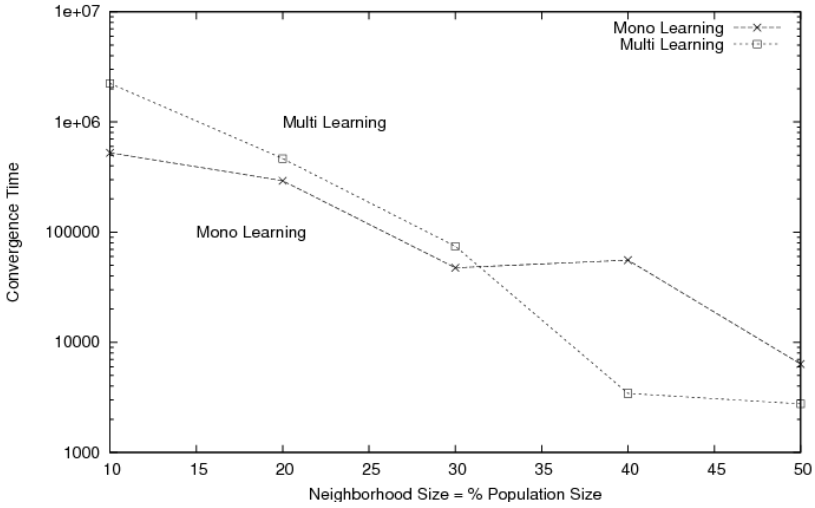


Figure 3.7: Different Learning Approaches in One Dimensional Lattices with different Neighborhood Sizes

conflicts, relatively more interactions between the agents in the overlap area between regions adopting conflicting subconventions is necessary. Unfortunately, agents in the overlapping regions may have more connections in their own subconvention region and hence will be reinforced more often by their subconventions, which makes it harder to break subconventions and arrive at a consistent, uniform convention over the entire society. In the case of the mono-learning approach, the agents in the overlapping region will not be disproportionately reinforced by the other agents sharing its subconvention, making it easier to break those subconventions.

On the other hand, when neighborhood sizes are large, and hence network diameters are small, agents interact with a larger portion of the population. This makes it more difficult to create or sustain subconventions. In addition, this large neighborhood size is more effectively utilized by the multi-learning as agents will be learning from all the interactions they are involved in, and not only from the interactions initiated by them.

For the scale-free and fully-connected stars, systematic variation of neighborhood size is not possible in general. We do observe an interesting phenomenon for these kind of networks. When the multi-learning approach is used in scale-free networks and fully connected stars, subconventions are persistent and the entire population does not converge to a single convention! This is the first time in all of our research on norm emergence that we observed the coexistence of stable subconventions.

The explanation of this rather interesting phenomena can be found in the combination of the memory-based reward function and the inherent topologies of such networks. We present, in Figure 3.8, a portion of a representative Scale Free or Fully Connected Stars network where subconventions have formed. We see that agent 1 (hub node 1) and its connected leave nodes (nodes 10, 11, and 12) have converged to one subconvention

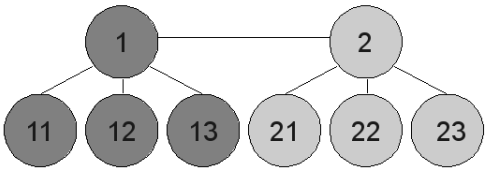


Figure 3.8: Subnetwork topology resistent to subconventions in a multi learning approach

(represented by the color of the nodes) that is different from the subconvention reached by agent 2 (hub node 2) and its connected leave nodes (nodes 21, 22, and 23). As an agent has equal probability of interaction with any of its neighbors, both agents 1 and 2 interact more frequently with their associated leave nodes that share their subconvention.

Also note that when two agents interact and both actions have been used equally often in their combined memories, as will be the case when agents 1 and 2 interact, the majority action will be selected randomly, giving advantage to one of the agents. For the subconventions to be broken in this scenario, it is needed that for one of the hub nodes the following holds true: (1) the agent's q-value for its preferred action decreases, and (2) the q-value for the action preferred by the other agent increases. In order for *the agent's q-value for its preferred action to decrease*, a number of repeated interactions (proportional to the memory size) between the hub nodes (in our example 1 and 2) have to occur, and as there will be no clear majority action, the preference has to be given to the same action, e.g., that preferred by agent 2, in all those interactions. As the reward for the agent 1's action will then be 0, its q-value will start decreasing. In order for the agent 1's q-value for its non-preferred action to increase, a number of interactions (also proportional to the memory size) between it and agent 2 has to occur and agent 1 has to explore in that interaction and try agent 2's preferred action. This will result in agent 1's estimate of agent 2's preferred action to increase, albeit slowly. Only when both these fortuitous events follow each other, and without the intervention of another interaction with the leaves associated with agent 1 (which would reinforce the subconvention), can the subconvention be ultimately broken. The likelihood of these sequence of events happening is exceedingly small and hence subconventions routinely arise with the multi-learning approach. Viewed another way, the leaf nodes can only interact with their hubs and each of them will reinforce the subconvention action for their associated hub node in every time step, making it very unlikely that conflicting subconventions will be resolved in situations such as in Figure 3.8.

On the other hand, as an agent is reinforced only once each time-step in the mono-learning approach, the processes required to break the subconventions are more likely, even though it sill has a relatively small probability. This probability decreases with larger memory sizes, and hence subconventions are more likely to emerge with larger memory sizes when using mono-learning. Therefore, with larger memory sizes, subconventions will be harder to break and this phenomenon caused the significant increase

in convergence time for scale-free networks and full-connected star networks (we discussed this in the previous section with reference to results displayed in Figure 3.3).

3.3.5 Effect of Number of Players

To investigate the effect of the number of players per interaction on the convention emergence process we designed the following experiment set. Agents are situated in a fully connected network. As the actual design of the Memory Based Reward Function does not consider situations with more than two players, we modify it to produce a new reward function (shown in Algorithm 3) that reinforces the convention taken by the majority of the players.

```
   // First, we select the majority action
1  Choose MajorityAction as the most frequent action over memories of all
   interacting players;
2  In case of tie, select randomly amongst the possible ones;
   // Then, we calculate the reward depending on the agents
      action selection and on the majority action
3  if Action_1 == MajorityAction then
4  |  reward_1 = 1
5  else
6  |  reward_1 = 0
7  end
```

Algorithm 3: MultiPlayer Reward Function.

Results from this experiment are presented in Fig. 3.9. We observe that the convergence time seems to decrease when the number of players in each interaction increases. The reason of this phenomena is found on the design of the reward function: agents will reinforce the action played by the majority. In situations with smaller number of players, fewer agents are consulted while deciding the majority action. This leads to the emergence of more subconventions, and these agents reinforce each other. On the other hand, in scenarios with larger number of players per interaction, the majority action will be calculated by consulting the history of more agents, and therefore, these subconventions will be less likely to appear. In essence, more consistent rewards are provided to a larger number of agents, hence promoting convergence of choices and enabling the emergence of a global convention. Information sharing between a larger number of individuals then have a similar effect on global convention emergence as the likelihood of interacting with a larger proportion of the society. Both facilitate uniform action adoption.

In Figure 3.10 we can observe how the neighborhood size affects (as identified in Section 3.3.3) the emergence of multiplayer conventions. We previously observed that the emergence of conventions in one-dimensional lattices was strongly affected by the neighborhood size: the convergence times are drastically reduced when the diameter of the network (inversely proportional to the neighborhood size) is reduced. We can now

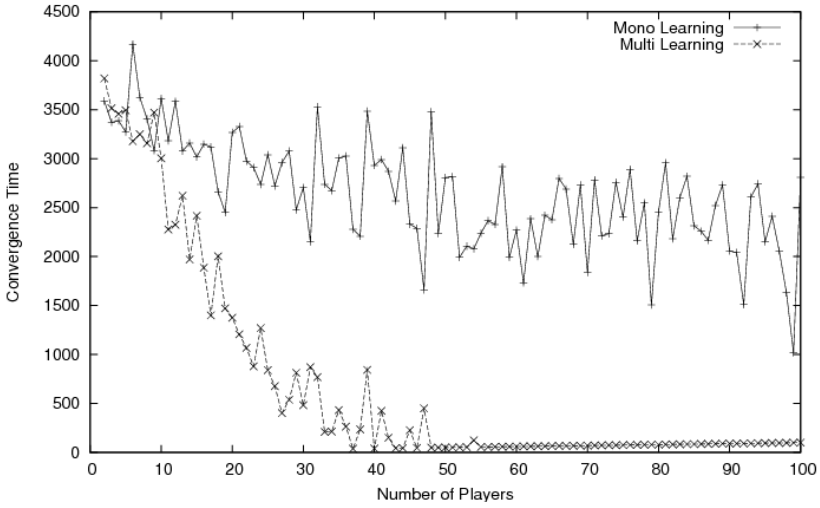


Figure 3.9: Different Players in Fully Connected Networks with Different Learning Approaches

observe similar effects in Multiplayer situations. We also observe that the convergence time also increases when increasing the number of players. This phenomena occurs due to the learning modality used by the agents (a multi-learning approach) and the relation between the number of players and the neighborhood size. When a small number of players interact (relative to the neighborhood size), subconventions are less likely to be created. However, when the number of players per interaction is larger (and closer to the neighborhood size), agents will always interact with the same agents, creating and reinforcing subconventions.

The neighborhood size also plays an important role when dealing with multiplayer interactions. Assuming that all the players in an interaction must be neighbors, we can infer that the amount of neighbors an agent has directly affects to the convergence time. We have accordingly designed two methods that will adapt the neighborhood size of the network, depending on the number of players per interaction: the reduced and the super-reduced. The *super-reduced* scheme assigns the minimum amount of neighbors needed for the number of players specified. Recall that the neighborhood size should be an even number N, representing $N/2$ of neighboring agents on each side of the agent. Therefore, in the super-reduced method, for either a 2 or 3 players games, the neighborhood size will be $N = 2$. We have observed that using the *super-reduced* method, in the case of the upper limit of players (3 players with a neighborhood size $N = 2$), makes agents interact repeatedly with the same agents every timestep. Depending on the type of reward function, this repeated interaction with the same players would lead to metastable subconventions. Therefore, we relaxed this method creating the *reduced* method. The *reduced* method assigns the minimum amount of neighbors needed for

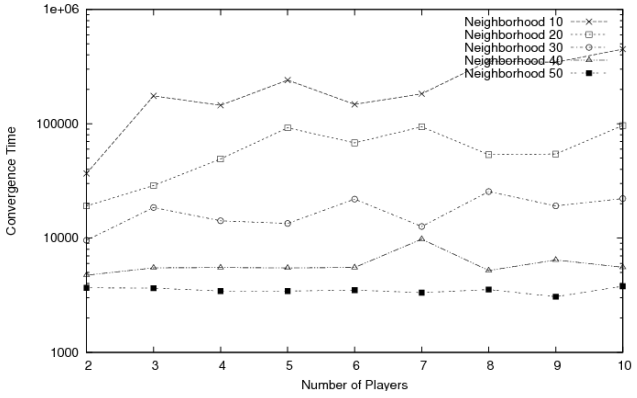


Figure 3.10: Neighborhood Sizes Comparison with Different Players with Multi Learning Approach

the amount of players specified *plus 2*. This addition of two extra neighbors introduces variety into the games (reducing the endogamy), allowing agents to play with different agents in different interactions.

From the results presented in Figure 3.11 we can observe the convergence times of both methods (reduced and super-reduced) with different number of players per interaction. We can observe how the *super-reduced* method produced larger convergence times and more pronounced changes between the even and odd value of each neighborhood size[2]. The reason why the *super-reduced* takes longer to converge is because of the interacting topology. When agents are not allowed to interact with other agents than their direct neighbors, no variance is introduced (complete endogamy), promoting the appearance of a "frontier effect". This frontier effect affects the agent located in the middle of two regions of clear preference: this agent in the "frontier" will be doubtful about its preference (one example of this situation can be seen in Fig. 3.12). If the agents that the frontier agent interact with are always the same, it will be harder to break the frontier effect this agent is under the influence of.

Until one of the involved agents explores a different action, this frontier effect will not be broken.

Moreover, in Fig. 3.11 we observed another phenomenon that has not been explained yet. We observe that the scenarios with an odd number of players games take longer to converge than those with an even number of players. We can find an explanation for this phenomenon in the design of the reward function and the frontier effect: the majority action is chosen by observing the actions of the interacting agents, and, if those are even, there might be a tie. In case of a tie, the majority action is chosen randomly, giving a clear advantage to one of the actions in the "frontier" region. In the case of an odd amount of players, the frontier agent will be affected by the same amount

[2]For $N = 2$, 2 and 3 players games can be played. For $N = 4$, 4 and 5 players games can be played. For $N = m$, m (even) and $m+1$ (odd) players games can be played.

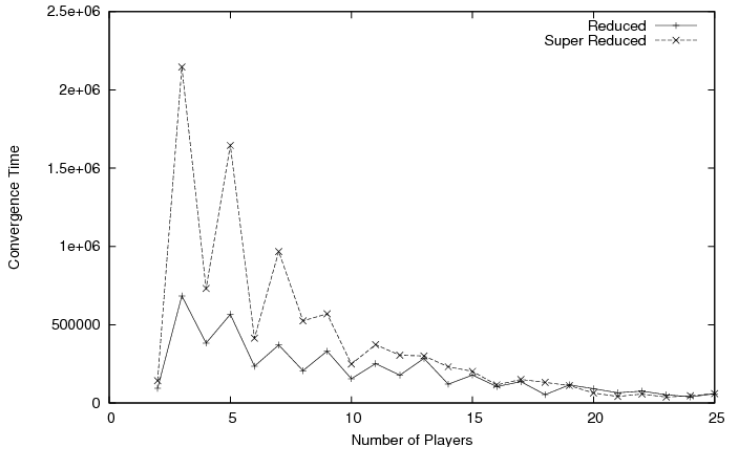


Figure 3.11: Reduced and Super Reduced Neighborhood sizes with different Players using the Multi Learning modality.

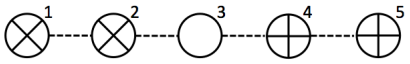
of non-frontier agents. We can observe and example scenario in Figure 3.12. Initially, in Figure 3.12(a), Agent 3 is doubtful about its preference. When Agent 2 interacts, in Figure 3.12(b), it affects (together with agent 1) agent 3, biasing their action preference to one convention. On the other hand, when agent 4 interacts, it will affect (together with agent 5) agent 3, making it change its preference to the other convention.

The random selection of the majority action in case of a tie, possible only for even number of players, speeds up the process of convergence, although producing a larger number of preference change per agent.

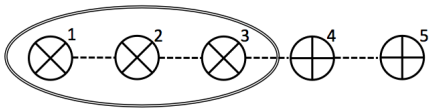
On the other hand, when having an odd number of players, we obtain the previously explained "frontier" effect. This effect ensures a longer convergence time (because the frontier agent needs to explore in order to break the frontier) with a smaller number of preference changes. Experimental results are shown in Figure 3.13 confirming our hypotheses.

3.3.6 Effect of Action Set

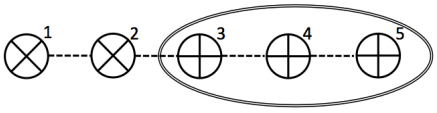
This last experiment was designed to evaluate how the topology plays an important role in the emergence of conventions in environments where there are more than two conventions to choose from. We can observe in Figure 3.13, that the larger the search space in the number of possible conventions, the longer it takes for the conventions to emerge. Specially pronounced is the effect with smaller neighborhood sizes. The explanation for this phenomena is again the creation of metastable subconventions promoted by the endogamy produced by the topological structure as explained in Section 3.3.3. However in the previous experiments, agent only had two options where to choose from. Having a larger set of options generate a larger number of different subconventions that need to be broken.



(a) Agent 3 doubtful



(b) Agent 2 interacts and affects Agent 3



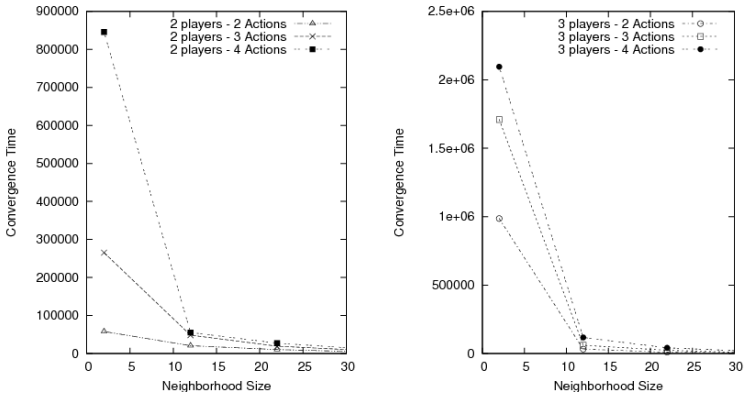
(c) Agent 4 interacts and affects Agent 3

Figure 3.12: Frontier effect in 3-players game with two competing conventions.

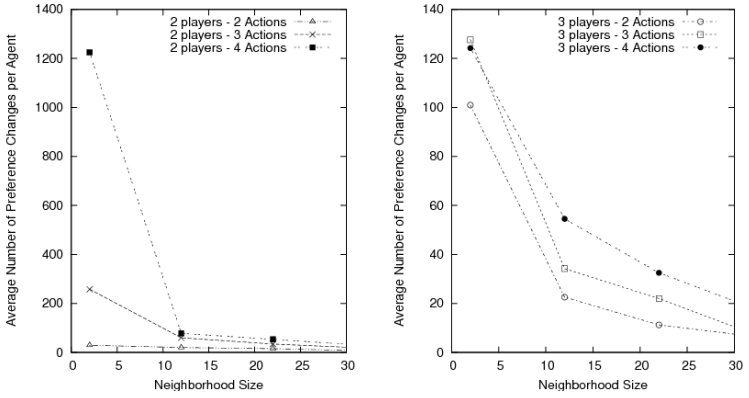
3.4 Discovering subconventions: not only history, but topology

These first set of experiments have been useful to understand the emergence of social conventions based not only on direct interactions but also on the memory (and previous history) of each of the agents under different interconnection topologies between agents. Our initial hypothesis was that subconventions are facilitated by the topological configuration of the environment or by the agent reward function. Experimental results using the *Memory Based Reward* function confirmed that the cultural maintenance promoted by this functions slows down the convention emergence and in certain situations provoke subconventions to emerge.

However, empirical results have given us a number of proofs of the effect of topology on the maintenance of subconventions. Specific topology structures promote endogamy between a certain subgroup of agents that might produce a different subconvention than the rest, as identified by several authors [Salazar-Ramirez et al., 2008, Toivonen et al., 2009, Villatoro et al., 2009]. Specially, we have observed that subconventions are more likely to appear and are more resistant when using the multi-learning approach, and might not be resolved for scale-free and fully-connected star networks. This is an important result as we consider the multi-learning approach to be the most realistic. From now and on in the remaining of this thesis we will only consider that learning approach.



(a) Convergence Time for 2 Player Games (b) Convergence Time for 3 Player Games



(c) Average Number of Preference Change for 2 Player Games (d) Average Number of Preference Change for 3 Player Games

Figure 3.13: Multi Player and Multi Action Games

Chapter 4

Unraveling Subconventions

The problem of subconventions is a critical bottleneck that can derail emergence of conventions in agent societies and mechanisms need to be developed that can alleviate this problem. Subconventions are conventions adopted by a subset of agents in a social network who have converged to a different convention than the majority of the population. As we have seen, subconventions are facilitated by the topological configuration of the environment (isolated areas of the graph which promote endogamy) or by the agent reward function (concordance with previous history, promoting cultural maintenance). Assuming that agents cannot modify their own reward functions, the problem of subconventions has to be solved through the topological reconfiguration of the environment.

4.1 Social Instruments for Subconvention Dissolution

As part of the social network, agents can exercise certain control over their social network so as to improve one's own utility or social status. We define *Social Instruments* to be a set of tools available to agents to be used within a society to influence, directly or indirectly, the behaviour of its members by exploiting the structure of the social network. Social instruments are used independently (an agent do not need any other agent to use a social instrument) and have an aggregated global effect (the more agents use the social instrument, the stronger the effect).

Normally, in *social learning* scenarios the reward function for an agent is determined either by the environment (e.g., if you over-exploit resources, they will be exhausted) or by individual social interactions (e.g., drive on the same side of the road as the rest of the drivers for your and others' safety) and typically individual agents cannot modify it. However, in *social learning* interactions, the reward is determined based on the strategies of the partners involved in the interaction. Therefore, an agent can indirectly modify the reward obtained and change its success and visibility and hence social status by controlling or modifying the social network it is located in, e.g. selectively choosing which other agents it interacts with, giving different importance to the relationships with other agents, share information about the interactions with other agents,

etc.

Social instruments are used by agents to improve their utility, without directly altering the utility function. Given a set of social instruments, agents can exploit the state of their social environment to have tangible effects on the utility obtained from their interactions with other agents in that environment.

A society where individual agents repeatedly use a social instrument has the properties of a complex system: the effect of the individual elements aggregated together exhibit, as a whole, properties that are not present in the properties of the individual parts. Moreover, the emergent aggregate behavior may change over time. The impact of the social instruments is reflected on its aggregated effect which often transforms the social environment.

For source of inspiration, we can look to human societies and observe how humans have developed different social instruments to exploit and produce changes in their social networks. These human social mechanisms have direct or side-effects on the social networks. There exists some literature on classifying and analyzing the effects of these social instruments on the network and therefore on the agents' interactions [Albert et al., 2000, de Pinninck et al., 2007b, Savarimuthu et al., 2007a, Urbano et al., 2009a, Villatoro et al., 2009, Babaoglu and Jelasity, 2008, Conte and Paolucci, 2002, Hales, 2002, Hales and Arteconi, 2006].

However, there is a critical dearth of work in this area of socially-inspired tools for multi-agent systems (MAS). Our goal is to fill this gap by formalizing and adapting the social mechanisms used by humans for agent societies. As a social network can be represented as a type of graph, graph theory provides us with a good ground formalism and allows us to establish a parallel between network operators and social operators. However, from a societal point of view only a subset of all the possible graph operations make sense.

In a social network, agents are represented by nodes and relationships by edges. Assuming agents cannot create or destroy other agents, the set of graph operations that are socially relevant are the following: destroy a link, create a link, add attributes to nodes (node coloring), add attributes to links (link coloring) and transmit information (any node can transmit information to any other node following the right protocols using the communication network, which might be different from the social network).

The social instruments available to the agents should be able to perform those operations. To alter the state of the social network in which agents are located, researchers have developed different social mechanisms that we now categorize. We identify some social instruments found in the MAS literature and observe their effect on the social network. The mechanisms we are interested in are those that can be used by the agents without institutional support.

- **Partner Selection (Edge Removal)**: People continually select who they do and do not want to interact with. Giving agents the capability of removing edges attached to them in their social network, gives them direct control over the network that directly affects them, improving the overall behavior of the system [Albert et al., 2000, de Pinninck et al., 2007b].

- **Tags (or Node Coloring)**: Recognizing certain attributes and characteristics of other agents before interacting with them reduces social friction (e.g., by reducing the number of unsuccessful interactions) and improves coordination. To facilitate this process, some kind of externally visible social markings are needed, e.g., tag mechanisms [Chao et al., 2008, Holland, 1993]. By carrying a tag, having a role [Savarimuthu et al., 2007a] or a certain *force* [Urbano et al., 2009a], an agent allows all other agents to recognize a certain characteristic or quality even before any direct interaction.

- **Social Position (or Link Coloring)**: Humans recognize that our behaviour has to be adapted depending on who we are interacting with, e.g., I behave differently with a friend from school than with my boss. Agents can extract information from the topology of a social network to gauge the influence or status of other agents and use this information strategically to further their own interests [Villatoro et al., 2009].

- **Mimicking (or Node Imitation)**: Humans consciously and sub-consciously imitate the strategies of more successful individuals as they aspire to improve their own situation. Access to the strategies of other nodes in the network can be used as a social instrument [Hales and Arteconi, 2006]. Such access can be applied to compare the efficiencies of several strategies and mimic the most effective ones.

4.1.1 Our Social Equipment

Some researchers in the literature have used certain instruments for the establishment and distribution of norms in virtual societies. However we have envisioned two new instruments that we hypothesize will act efficiently on the dissolution of subconventions: rewiring and observation. As we will see in the next sections, rewiring directly changes the structure of the network and observation exploits the information of certain parts of the network to bias agents' decisions.

Rewiring (or Intelligent Link Removal and Creation)

If we use human societies as inspiration, agents should have the ability of choosing whom they want to interact with. We have adapted this idea into a social instrument that allows agents to "break" the relationships from which they are not receiving any benefit and try to create new ones. This social instrument allows an agent to remove links with other agents it is connected with, and substitute intelligently those links by new ones, making this last step the crucial difference with *edge removal*. Agents decide to rewire a link after the number of unsuccessful interactions[1] with another agent goes above a certain *Tolerance* threshold. Agents also need to decide whom they want to establish the new link with. We have developed three different methods:

[1]Unsuccessful interaction in our convention emergence scenario corresponds to being uncoordinated or not sharing the same convention for that interaction.

1. *Random Rewiring*: Agents rewire to a randomly selected agent from the population.

2. *Neighbour's Advice*: Agents rewire to an agent recommended by a neighbour. The rewiring agent, X, asks a neighbour with the same preference as that of X (agent referent) for another agent also with the same preference but not a neighbour of X (agent referred). If it was not possible to achieve that, Random Rewiring will be applied.

3. *Global Advice*: Agents rewire to an agent that is randomly selected by the system from those that have the same strategy. This rewiring technique is sued as control case.

Despite the similarity with the work of Griffiths [Griffiths and Luck, 2010], there exist a crucial difference with our approach: both [Griffiths and Luck, 2010] use an evolutionary approach, observing the results of their techniques after the reproduction of a number of generations, and with a certain mutation rate. On the other hand, we use a more *online approach* where agents can modify their social network on runtime, without the necessity of evolving new generations. In addition, our rewiring methods do not access any private agent's information (used only in the Global Advice which is used as a control case), such as their actual reward. There is also a clear difference of this social weapon with the *partner selection* mechanism: rewiring produces a tangible change in the network, partner selection on the other hand does not directly affect the network but prioritise some nodes over the rest.

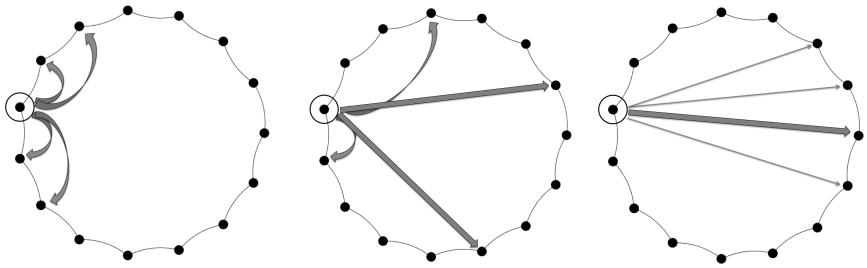
Observation

In a social learning scenario, allowing agents to observe the strategy of other agents outside their circle of interaction can provide useful information to support the convention emergence process. However, there has to be a trade-off between observing and interacting. Allowing agents to observe other agents' state might help when deciding which strategy to adopt. However, there has to be a trade-off between observing and interacting. In order to analyze the effects of observation we will allow agents to observe, at certain timesteps, a subset of agents' states in the population. Therefore, agents will be assigned an *Observation Probability*. Moreover, agents need to know the amount of agents they can observe (*Observation Limit*) and how they want to observe (*Observation Method*). We propose three different observation methods:

1. *Random Observation*: Agents observe random agents from the society. (Figure 4.1(b)).

2. *Local Observation*: Agents observe their immediate neighbours in the social network. (Figure 4.1(a)).

3. *Random Focal Observation*: Agents select one random agent from the society and observe that agent and its direct neighbours. (Figure 4.1(c)).



(a) Local (b) Random (c) Random Focal

Figure 4.1: Observation Methods

After the observation process, the agent will choose the majority action taken by the selected observed agents and will reinforce it. We are aware that in certain configurations, choosing the observed "majority action" might not help agents to achieve the general convention. Nonetheless, and because of the intrinsic nature of the convention emergence process, the observed "majority action" will lead in most of the situations to the general majority action.

Despite the similarity, this instrument and *mimicking* (presented in the previous section) behave differently. Mimicking agents copy from a successful agent the strategy (which in most of the cases is private) as well as the interaction neighbours (which might also be private). On the contrary, with the observation instrument agents observe the last decision (hoping this represents the actual strategy) of a group of agents (without knowing the actual success of the agents) and update their estimates based on the majority. With observation, agents only access information that has been previously made public by the observed agent, while with *mimicking*, they access information that can be considered private.

4.2 Experiments

We use the simulation model presented in Section 3.1 to evaluate how the social instruments that we propose influence the process of emergence of social conventions.

However, and in order to obtain complete results, we have introduced at this point different decision making functions to evaluate the efficiency of our social instruments. The decision making functions are:

1. *Best Response Rule (BRR)* [Mukherjee et al., 2007, Sen and Airiau, 2007]: Agents choose the action with which they have obtained the highest payoff in the last iteration. A positive reward is given to agents if they are coordinated. This function gives preference to the instantaneously utility maximizing action.

2. *Highest Cumulative Reward Rule (HCRR)* [Shoham and Tennenholtz, 1997a, Kittock, 1993]: "an agent switches to a new action iff the total payoff obtained

from that action in the latest *m* iterations is greater than the payoff obtained from the currently-chosen action in the same time period". A reward is generated based on a coordination game. This function gives preference to the action that has obtained the higher accumulated payoff in the last *m* interactions (rather than only in the last one).

3. *Memory Based Rule (MBR)* [Villatoro et al., 2009]: A positive reward is given to agents if they are coordinated, and it is proportional to the actions they have chosen in the past. This function gives preference to the action that has provided the largest payoff while taking into consideration in the reward function also the previous actions (promoting concordance with previous history).

For all the following experiments we fix the population size to 100 agents (as we have been able to understand the different population size effects so far) amd the memory size to 5 (for HCRR and MBR).

4.2.1 Rewiring

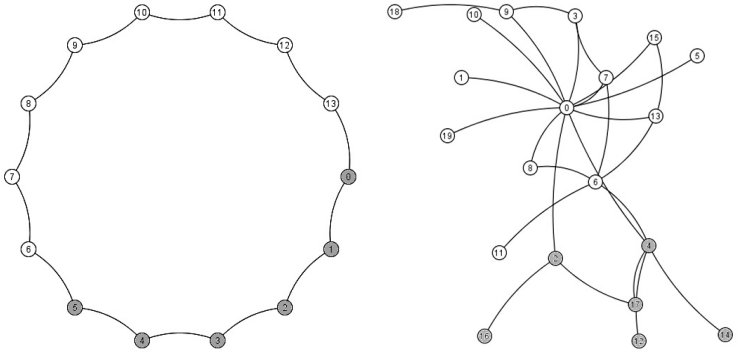
We have experimented with the three rewiring methods introduced in Section 4.1.1 on three topologies: a low clustered [2] one dimensional lattice (lattice with Neighborhood Size = 10), a high clustered one dimensional lattice (lattice with Neighborhood Size = 30), and a scale free network. We have explored the search space of the Tolerance Levels, observing how they affect the convergence time and the number of components (maximally connected subgraphs) created when convergence is reached with the different decision making functions.

Influence of Rewiring Methods

From the experimental results obtained we have observed how, in general, the *Global Advice (GA)* rewiring method produces the best convergence time due to its centralized nature and access to global information. Nonetheless the decentralized methods, specially the *Neighbor's Advice (NA)* method, also show good performances. The *NA* method improves the *Random Rewiring (RR)* method as it more expediently resolves the subconventions that appear in the one-dimensional lattices during the convention emergence process. These metastable subconventions have been already identified in [Salazar-Ramirez et al., 2008, Toivonen et al., 2009, Villatoro et al., 2009] to be the preferred action of a group of nodes who have converged to a different convention than the rest. These subconventions, as noted in [Villatoro et al., 2009], do persist as long as the frontier region remains metastable[3]. The frontier region is the group of nodes in the subconvention that directly interacts with other nodes that have adopted a different convention (if nodes represent agents, and the color of the node the convention they have adopted, Fig 4.2 shows examples of frontiers in two different networks).

[2] Clustering Coefficient is a measure of degree to which nodes in a graph tend to cluster together.

[3] Metastable conventions are those that should preferably be broken, but remain stable because of a combination of factors (like interaction dynamics, topology and decision making functions).



(a) Frontier in Ring Network. (b) Frontier in Scale Free

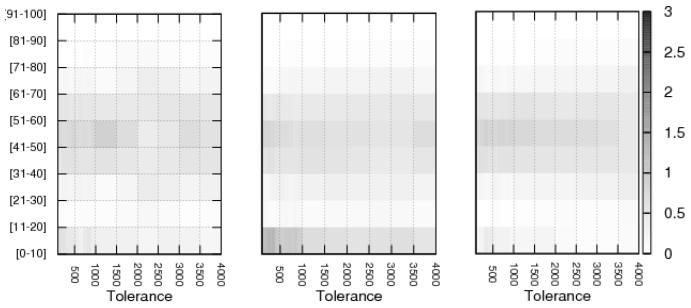
Figure 4.2: Examples of Frontiers in Different Networks.

When using the *Neighbor's Advice* method, these subconventions are resolved more expediently. Agents in the frontier use the rewiring instrument as they cross the tolerance level faster than those not in the frontier. For this reason, the *RR* method will relink an agent with a more suitable agent with a probability of $\frac{1}{NumberOfActions}$. In contrast, the *NA* method will relink the agent with another one with the same preference if it is accessible. In case there is no other agent with the same preference to connect with, random rewiring will be applied, obtaining in the worst case scenario, the same results.

These results are reaffirmed for the scale-free networks, although an interesting phenomenon concerning the final number of components is observed in this case. Figure 4.3 shows the size of the clusters when a convention was reached. The horizontal axis represents the Tolerance levels and the vertical axis presents ranges (of size 10) of cluster sizes, and the histogram darkness the amount of clusters. When using the *RR* (Fig. 4.3(a)) and the *GA* (Fig. 4.3(c)) methods, the number of components obtained is always above one, resulting in a fragmented society. The number of components is slightly higher for the *NA* method than when using the other methods. This phenomenon is observed because of the structure of the scale-free networks (very few nodes with high degree (hubs) and a larger amount of nodes with a lower degree following a power-law distribution) and the dynamics of the rewiring methods. Non-hub nodes reach their tolerance levels faster than hub nodes (as non-hub nodes interact with a lower amount of agents). Once the tolerance level is reached, the rewiring process starts working. We do not observe any distinctive behavior separating the *RR* or the *GA* methods as they both incorporate a random component. However, when using the *NA* the number of components is variable, depending on the tolerance level: with a low tolerance the number of components is higher, although we have observed that most of them are smaller components (as it is shown in Fig. 4.3(b) observing larger amounts, represented with darker colours in the figure, in the smaller cluster ranges).

Non-hubs (when using the NA method) will rewire to any of their leaf nodes[4] sep-

[4]A Leaf node is a vertex of degree 1



(a) RR Method (b) NA Method (c) GA Method

Figure 4.3: Cluster Sizes Histogram for Scale Free

arating completely from the initial complete component (e.g. in Fig. 4.4), creating a greater number of components, but with smaller size.

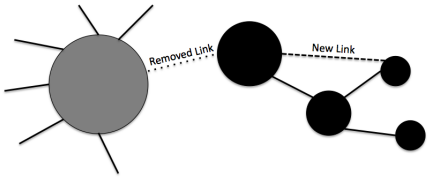
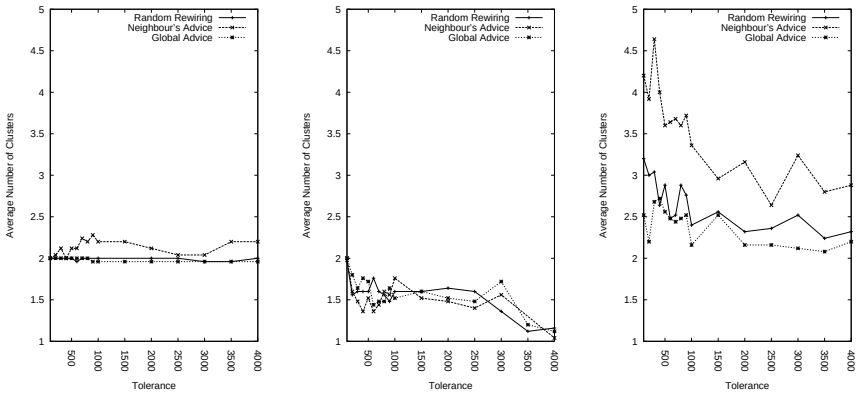


Figure 4.4: Example of Components Evolution in Scale Free NA Rewiring Method.

Influence of topology

When observing the effects of the topology, we find that the convergence time is increased under the effects of rewiring when the neighborhood size is increased. This effect is due to the clustering coefficient of the network. The one dimensional lattices with higher neighborhood sizes are more fragmented than those with more restricted neighborhoods. Therefore, when increasing the neighborhood size, the number of links between agents also increase, thereby increasing the clustering coefficient. Highly clustered societies are more resistant to rewiring, as the node that wants to use the rewiring would have to apply it to a higher number of nodes, and then, be rewired to the same amount of nodes with the appropriate strategy. This is one of the scenarios where the difference between our social instrument and that presented in [Hales and Arteconi, 2006, Griffiths and Luck, 2010]: our instrument needs to substitute one by one the selected links; on the other hand, in [Hales and Arteconi, 2006,



(a) NS = 10 (b) NS = 30 (c) Scale Free

Figure 4.5: Rewiring Methods Comparison for Number of components in a One Dimensional Lattice

Griffiths and Luck, 2010] their instrument accesses the private information of another agent and copies their set of neighbors. Our social instrument sacrifices convergence time for the sake of privacy.

Tightly linked to the previous results, we also observed that under the effects of rewiring there exists a trade-off between the diameter of the network and the number of components: with higher diameters (smaller neighborhood size) the number of components remains constant and greater than one (as it can be seen in Figure 4.5(a)). However, we can observe in Figure 4.5(b) that when the diameter is smaller (with higher neighborhood size), the number of components is reduced to one with certain values of *Tolerance*. The explanation of this phenomenon is again related to the clustering coefficient of the network combined with the dynamics of the rewiring process. When the clustering level of the network is low, agents have less links to be rewired. Initially, and before the *Tolerance* level is reached, agents also change their preferences without rewiring. Eventually, and after the *Tolerance* level is reached, agents start rewiring their links (and reducing their number of preference changes). As it was explained previously, low clustered societies have higher tendency to become disconnected and form different components.

Experimental results (shown in Figure 4.6) also show interesting properties with Scale Free networks (see Section 4.2.1). Similar to what was observed by [Villatoro et al., 2009] we found that subconventional effect has a stronger effect on Scale Free networks when using MBR, producing (as it can be seen in Figure 4.5(c)) generally a larger number of components. However, when using the rewiring social instrument, global convergence is obtained in this type of networks.

We can conclude that rewiring performs better in low clustered societies, producing a stratified population which results in a significant reduction in convergence time. In more clustered networks, the tolerance level has to be chosen carefully (depending on the other experimental parameters) to produce an effective technique for norm emer-

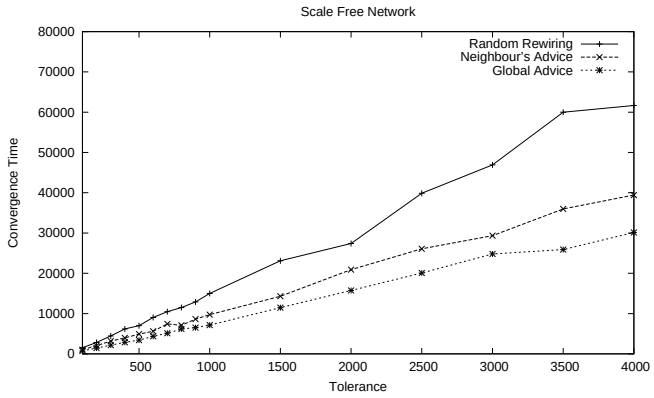


Figure 4.6: Convergence Times with different Rewiring Methods in a Scale Free Network.

gence.

4.2.2 Observation

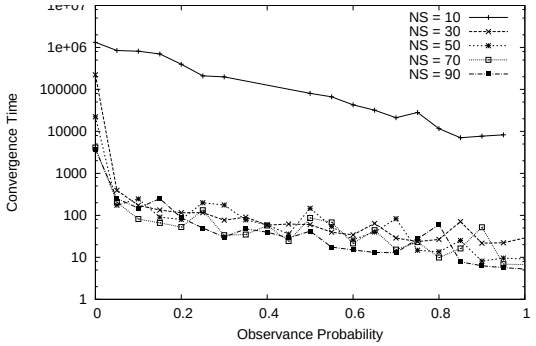
In this section we analyze the effects of *observation* as a social instrument when used by agents. We test and compare the three different methods proposed, exploring the search space with a representative range of *Observation Probability* values. As we did in the previous section, we experiment with three topologies: a low clustered lattice, a high clustered lattice and a scale free. To observe the effects of the different observation methods, we fix the *Observation Limit* to 10 for the experiments.

Influence of Observance Methods

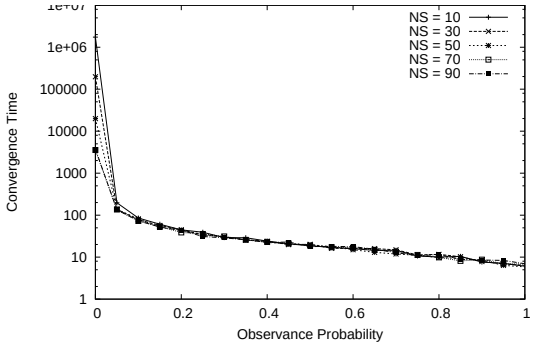
Comparing the results from the three Observation methods we observe in Figure 4.7 that the Random (*RO*) and the Random Focal Observation (*RFO*) methods are the most effective ones (with different topological effects to be commented), and have very similar results, when compared with the Local Observation (*LO*) method. The reason for this phenomenon is to be found on the frontier effect. When agents use the *LO* method, they observe their direct neighbors. If the observing agent is in the frontier area, then, this observation is pointless. However, observing different areas gives a better understanding of the state of the world, and hence the *RO* and the *RFO* methods perform better.

Influence of topology

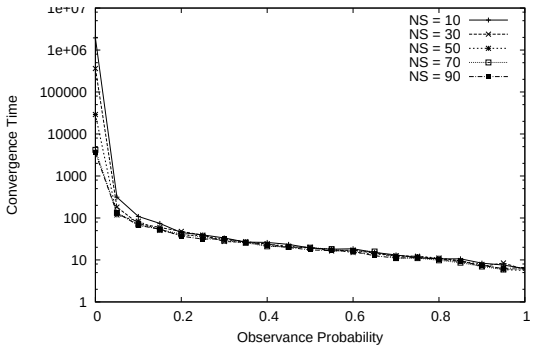
For the BRR and MBR decision making functions, we have observed that the different Observation methods produce a more pronounced effect in societies with higher diameters, as we can see in Fig. 4.7. We notice that a small percentage of Observation



(a) Local Observation



(b) Random Observation



(c) Random Focal Observation

Figure 4.7: Comparison on Effects of Convergence Time on One Dimensional Lattice.

drastically reduces convergence times. The reason for this effect can again be found in the frontier and the subconvention effect previously discussed. Subconventions emerge more readily when the social network has a small diameter and the frontier region represents the unsettled area. These subconventions are more easily resolved at these frontiers by observation rather than by learning through interactions.

Influence of Strategy Decision Techniques

The main experimental results obtained for BRR and MBR is that greater observation leads to faster convergence. However, in most cases, even smaller observation percentage helps significantly reduce the convergence time. This result is important if Observation has a concomitant cost. However, *observation* has a poor effect on HCRR: we find that the convergence time is proportional to the observation probability (the higher the observation, the worst it performs). The reason of this effect is found on the HCRR's lack of exploration policy. When observing, an action is reinforced, and this action might be different from the one that the agent has converged to within its environment. This new reinforcement destabilizes the convergence that was already obtained. This effect is not produced with the others reward rules as they have an exploration rate that quickly corrects this destabilization.

4.3 Discussion on the Frontier Effect

The experiments performed up to now have shown us how the emergence of social conventions can be facilitated with the usage of our proposed social instruments, achieving good results. However, some practical information has to be kept for the appropriate usage of them: on the one hand, tolerance levels have to be carefully chosen when using rewiring. On topologies with large diameters the social network might suffer important changes affecting the number of components. Moreover, we have observed an interesting phenomenon on Scale Free networks: without the usage of the social instruments full convergence was hard to obtain; with the rewiring instrument convergence is reached producing a larger amount of components when using the *Neighbour's Advice* method.

On the other hand, a small percentage of observation is very helpful in the achievement of full convergence. Interestingly, when using the *Local Observance* method, this improvement is not as with the other methods.

With these pieces of information we can hypothesize that there exist a special type of subconventions that cannot be resolved at a local level with the methods proposed, specially in Scale Free Networks.

4.4 Understanding Subconventions in Scale-Free Networks.

The results from experiments presented above, together with the observations narrated by other authors [Epstein, 2000, Toivonen et al., 2009, Villatoro et al., 2009], con-

vinced us that subconventions are problematic obstacles to the emergence of global conventions. These subconventions thrive (amongst other reasons) because of the topological structure of the network where they emerge.

Because of the inherent structure of the scale-free networks (nodes whose connections follow a power-law distribution), interesting properties of the network emerge in the dynamics of the convention emergence process. To have a more detailed idea of how the topology affect the process, we have performed a detailed study of its dynamics.

4.4.1 Who's the Strongest Node?

In previous works in the literature [Sen and Airiau, 2007, Mukherjee et al., 2008], some authors fixed the behaviour of a number of agents in the population to observe how their unchangeable behaviour would affect the emergence of a convention. Those experiments were performed in regular networks and the position of the fixed learners on the network was not really considered. Inspired by those works, we have decided to apply such technique in irregular networks, locating the fixed learners in different positions of the network.

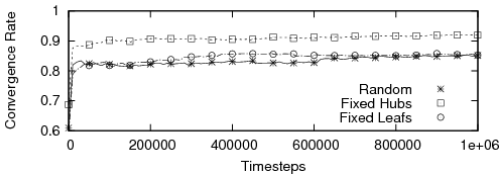
Scale free networks are basically defined as irregular networks where very few nodes have a large number of connections (hubs), and, a large number of nodes have a very few connections (leafs). In order to discover which nodes have a stronger effect in the emergence of conventions, we experiment by fixing a certain percentage of the population to an specific (and common to all of fixed-agents) convention. Three experimental variants are tested:

1. fixing hub agents' behaviour: the behavior of the nodes with higher degree will be fixed.
2. fixing leaf agents' behaviour: the behavior of the nodes with lower degree will be fixed.
3. fixing random agents' behaviour: the behavior of nodes selected randomly will be fixed.

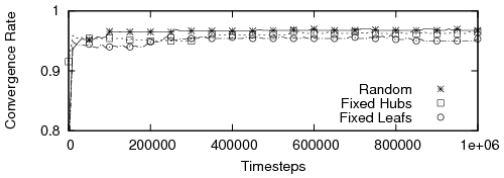
We hypothesize that by fixing the behaviour of hub agents, the rest of the population will converge faster than in any other possible situation. Our intuitions are founded on the dynamics of the convention emergence: given that a hub node will interact with more agents, its fixed behaviour will remain unaffected by the others' decisions. However, as others are learners, they will learn from the decisions of the hubs.

Experimental results presented in Fig. 4.8 partially confirm our hypotheses. We have performed an intense evaluation of the search space, exploring different percentages of fixed agents in different locations of the network and analyzing their convergence rate. For small numbers of fixed agents, fixing the hubs produces better convergence rates than when fixing the same amount of leafs or random agents (as we can observe in Fig. 4.8(a), Fig. 4.8(b), Fig. 4.8(c), and Fig. 4.8(d)). On the other hand, when the number of fixed agents is larger, fixing the leafs results in much faster convergence time (as it can be seen in Fig. 4.8(e) and Fig. 4.8(f)).

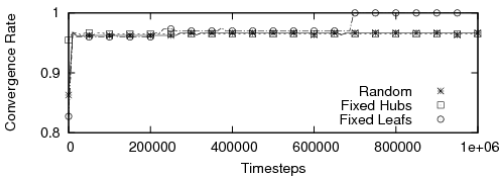
This empirical result lead us to pay attention on the leafs. Experimental data confirms that when the majority of the leafs have reached a common convention (artificially



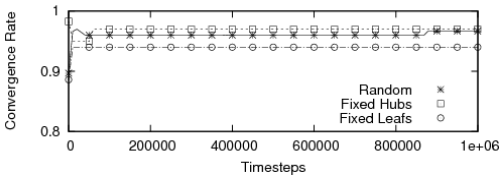
(a) 5% Fixed Learners



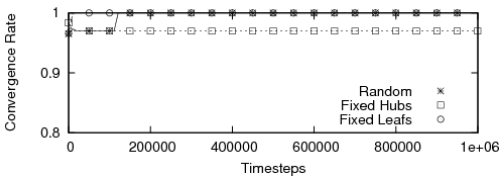
(b) 25% Fixed Learners



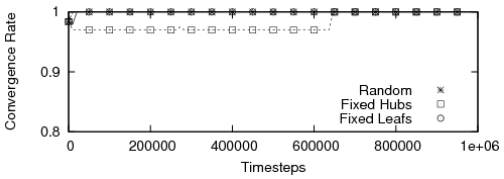
(c) 35% Fixed Learners



(d) 45% Fixed Learners



(e) 65% Fixed Learners



(f) 75% Fixed Learners

Figure 4.8: Fixing Agents Behavior

by fixing its behaviour, or naturally through the convention process) reaching the full convergence can be easily achieved. Consequently, in order to identify why learning leafs (by fixing the hubs' behavior) produces such delay in the convention emergence, we analyzed carefully the networks when the metastability was reached. By taking snapshots of the state of the convention emergence process (the structure and state of the network: which node is in which state and connected with which other nodes) at the moment where no improvement in the convergence rate has been made, we find what we have defined as *Self Reinforcing Substructures (SRS)*. These substructures are a group of nodes that, given the appropriate configuration of agents' preferences and network topology, do maintain subconventions.

During this research we have been able to abstract the topology of these structures into two general structures:

- The *Claw* SRS is formed by connecting a node with a number of *hangers*[5] connected to it smaller than the number of links with the rest of the network. In the situation where the hangers coordinate to the same convention among themselves and with the connecting node, we have a self-reinforcing structure. For example, in Fig. 4.9(a), A is the central node, having one connection with the rest of the network and 3 hangers: B (that it is another claw), C (plain hanger) and D (chain's connecting node).

- The *Caterpillar* SRS is a structure formed by a central path and from its members can hang other SRSs (such as claws, chains, or plain hangers). For example, in Fig. 4.9(b), A, B, C, and D are members of the central path, and the other nodes reinforce them.

These two abstract (examples in Fig. 4.9) structures can be found as subnetworks of scale-free and random networks. As we have observed, the existence of these SRS (74% of the generated networks with the methods described in [Delgado et al., 2003] contain SRS) are the main reason why convergence to a 90% level (as observed by [Delgado et al., 2003]) is achieved relatively quickly, but overcoming the last 10% (containing the SRS) is much harder to achieve.

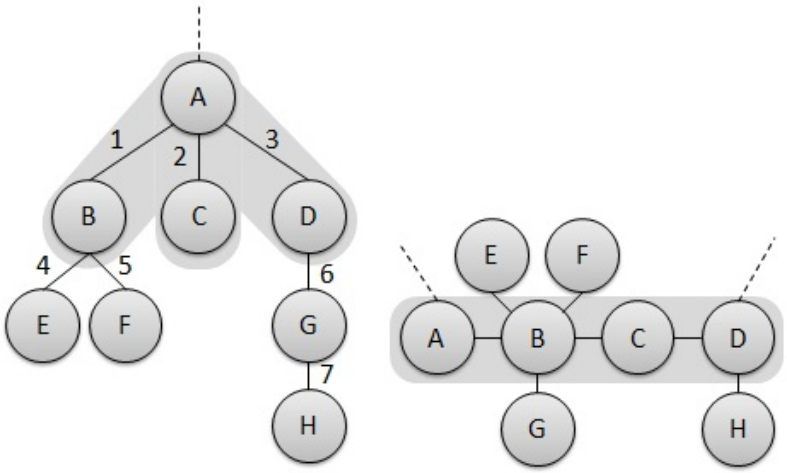
In order to check the validity of our hypotheses, we have extended the previous experiment, adding another variant where we fix the behavior of the nodes in the SRS with higher betweenness[6].

Experimental results in Fig.4.10 show a comparison of the results of convergence rates of the best performing variant of those presented previously and the one that fixes the agents' behavior in the SRS. We can observe that fixing the SRS nodes outperforms any of the other strategies.

This result is very interesting as it help us understand the specific substructures that generate subconventions within social networks. This result is applied in the next section to help us solve the subconventions effect in a more robust way.

[5] A *hanger* is formed by nodes that are connected to a member of a cyclic component, but which do not themselves lie on a cycle [Scott, 2000], and a *chain* is a walk in which all vertices and edges are distinct.

[6] Betweenness measures the extent to which a particular node lies "between" the various other nodes in the graph [Freeman, 1979].



(a) Claw (b) Caterpillar

Figure 4.9: Self-Reinforcing Structures

4.5 Combining Instruments: Solving the Frontier Effect.

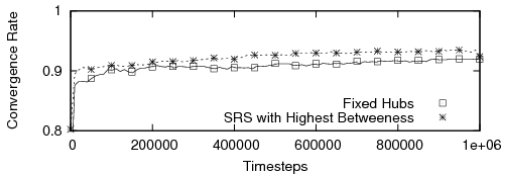
After experimenting with both social instruments in Sec.4.2, we observed that the subconventions need to be resolved in what we consider to be the "frontier" region.

Theoretically, a subconvention in a regular network is not metastable (i.e. the subconvention is continuosly affected tending to disappear), but unfortunately, slows down the process of emergence. On the other hand, in other network types, such as random or scale-free , subconventions seem to reach metastable states[7] because of the existence of the Self-Reinforcing Structures identified in Sec.4.4.1.

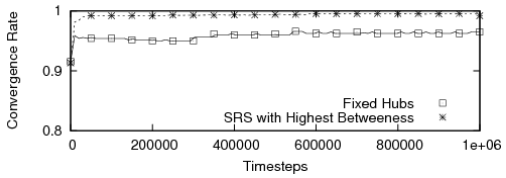
Therefore, by giving agents the tools to dissolve these frontiers, we hypothesize that convention emergence will be achieved faster and full convergence rates will be obtained.

However, because of the way social instruments have been designed, the instruments can be activated in situations where it is not necessary (observation follows a probabilistic approach, and rewiring is activated using a rewiring tolerance). Therefore, and by combining both social instruments developed in this work, we have designed a composed instrument for resolving subconventions in the frontier in an effective and robust manner. This composed instrument allows agents to "observe" when they are in a frontier, and then, apply rewiring, with the intention of breaking subconventions. To effectively use this combined approach, agents must first recognize when they are

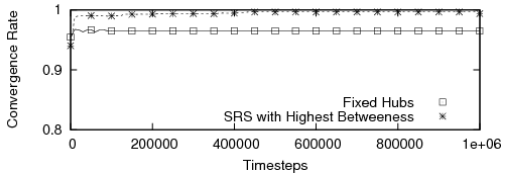
[7]By experimentation, we have observed that around 99% of the generated scale-free networks do not achieve full convergence before one million timesteps with any of the decision making functions used in this work and without any social instrument.



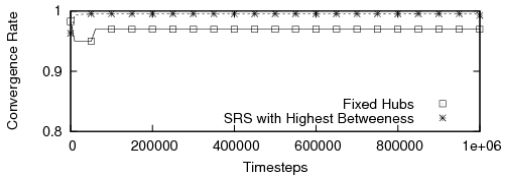
(a) 5% Fixed Learners



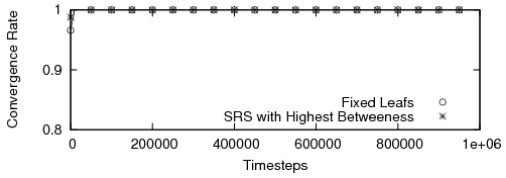
(b) 25% Fixed Learners



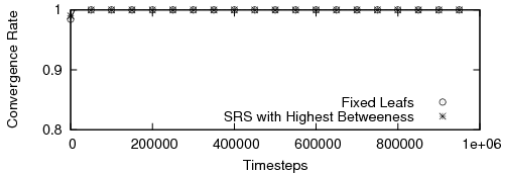
(c) 35% Fixed Learners



(d) 45% Fixed Learners



(e) 65% Fixed Learners



(f) 75% Fixed Learners

Figure 4.10: Comparison with Fixed SRS Nodes

located on a frontier.

We define a frontier as a group of nodes in the subconvention that are neighbours to other nodes with a different convention and at the same time are not in the frontier with any other group. To provide a more precise definition we first need to define our system:

$$System^t = \{A, Rel^t, S^t\} \tag{4.1}$$

where A is a set of agents, Rel^t is a neighbouring function at time t, S^t is the actual state of agents at time t.
We can therefore formalize the notion of subconvention *Sub*:

$$Sub^t \subset A \text{ where } \exists a \in Sub^t, \nexists b \in Sub^t \mid Rel^t(a,b) \wedge S^t_a \neq S^t_b$$

And now we can define when an agent is located in a frontier:

$$Frontier^t(a) = \{a, c \in Sub^t \wedge \exists b \notin Sub^t \wedge Rel^t(a,b) \wedge S^t_a \neq S^t_b \wedge Rel^t(a,c) \wedge (\forall d \in A \mid Rel^t(c,d) \rightarrow S^t_c == S^t_d)\} \tag{4.2}$$

This formula basically means that an agent a is in the frontier when it is in front of another agent (b) not sharing the same convention, and at the same time, is being reinforced by other agents (c) that are not in the frontier.

As the cases are different for regular or irregular networks, two types of frontiers need to be defined:

- *weak frontiers* as the ones that are not metastable in regular networks.
- *strong frontiers* as the ones generated by the SRSs in irregular networks.

The most important characteristic that defines a frontier is the existence of a confrontation. Confrontation occurs when two agents in an interaction do not share the same convention [8].

Before proceeding further, we will define three characteristics of agents with respect to their convention and the topological position in the network. An agent is *in equilibrium* if it has the same number of neighbours in its own convention as in the other convention. An agent is a *weak* node if the number of neighbours in its own convention is lower than those in the other, and an agent is a *strong* node otherwise (if the number of neighbours in its own convention is greater than those in the other). In regular networks, two confronted agents are in a frontier region iff: (1) At least one of the confronted agents is in an equilibrium position, and (2) all the neighbours of an in-equilibrium confronted agent are strong nodes. In irregular networks, two confronted agents are in a frontier region iff both agents are *strong* nodes.

With these formalizations and identification of what the frontiers are in each of the topologies, agents can be equipped with this knowledge and the tools necessary to recognize when they are located in a frontier and repair that situation. To achieve this

[8] Not sharing the same convention, choosing a different action, or choosing a different state to be, are considered equivalent expressions for our purpose.

task we will use the simple instruments presented in Sec. 4.1.1 but in a combined manner: firstly, agents will use *observation* (with the local observance method) to identify if they are located in a frontier (and therefore part of a self-reinforcing structure), and secondly, they will use *rewiring* to solve the frontier problem, by disconnecting from another agent in the SRS and reconnecting to another random agent.

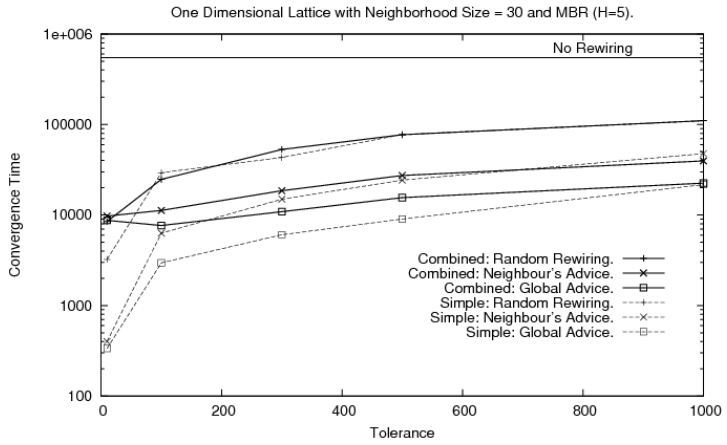
4.5.1 Results

We have conducted exhaustive experimentation with the composed instrument on the three topologies and using the different decision making functions described in the previous section. The use of the composed instrument on the regular networks does not produce an improvement on convergence time with respect to simple rewiring (one example of topology and strategy decision technique can be observed in Fig. 4.11(a)). However, an important improvement is observed in the number of rewired links (one example of this improvement can be seen in Fig. 4.11(b)). In general, this improvement is observed for lower tolerances. The reason of this effect is because for higher tolerances rewiring works in the same way as the composed social instrument, but without observing. For those smaller values, the effect is intense, reducing the number of rewiring links down to half of the original value.

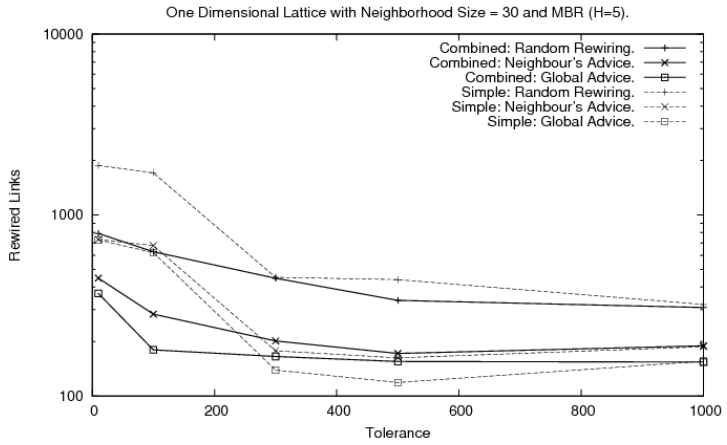
On the other hand we observe an important improvement for convergence times when using the composed instrument (with the recognition of SRS) on irregular networks. The results presented in Figure 4.12 represent the average results from 25 different scale-free networks with and without using the Combined Social instrument. By comparing Figure 4.12(a) and Figure 4.12(b) we notice the trade-off between the improvement in convergence time and the amount of rewiring to be done. The reason of this phenomena is because the Composed social instrument decomposes the SRS differently than the simple rewiring which only rewires the node in the actual frontier.

4.6 Conclusions

In this chapter we have taken a step forward in the state-of-the-art research on convention emergence by setting the convention emergence rate to 100%. This improvement might initially seem meaningless although our experimental results showed that overcoming the 90% rate is a challenging task in certain configurations of the environment because of the creation of subconventions. This research focused on identifying under what conditions subconventions emerge and are maintained along the time. We hypothesized that was mainly for two reasons: cultural maintenance and endogamy amongst the members. To prove the first we introduced a *Memory based reward rule* that simulated a function that promoted the cultural maintenance, analyzing carefully the situations where subconventions emerged. After a profound analysis we learnt that the subconventions are strongly affected by the topological configuration of the interaction network which promotes strong endogamy. As system designers we are still interested in achieving a 100% convergence rate (as it is the most efficient strategy for the system as a whole when dealing with conventions), and assuming agents cannot change their reward function, we have introduced the use of *Social Instruments* as tools that facilitate

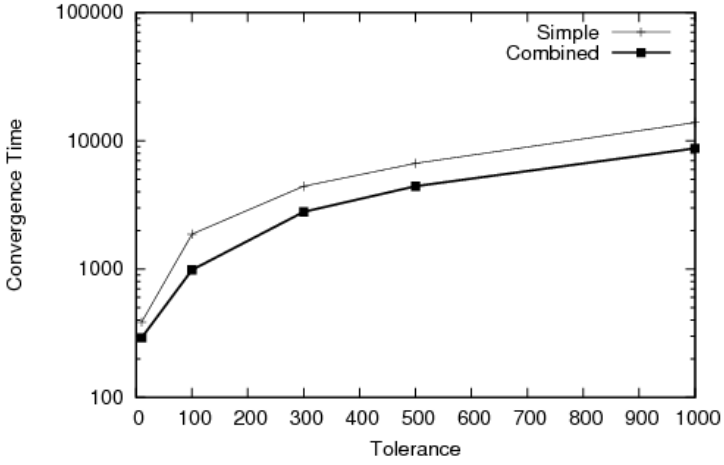


(a) Convergence Time

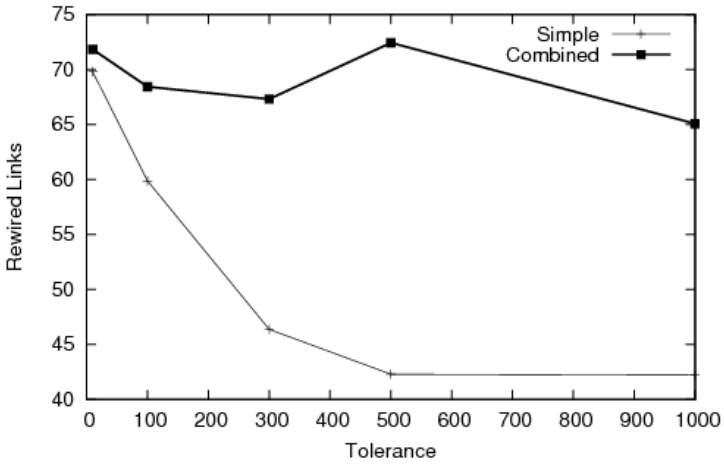


(b) Rewired Links

Figure 4.11: Comparison with Simple and Combined Social Instruments on Regular Network using MBR.



(a) Convergence Time



(b) Rewired Links

Figure 4.12: Comparison with Simple and Combined Social Instruments on Scale Free Network using BRR.

norm evolution. We have identified the characteristics and opportunities for effectively utilizing these social instruments for facilitating norm emergence through social learning. Social instruments are attractive since they do not require centralized monitoring or enforcement mechanisms, normally are extremely easy to use, have very low computational costs, and are scalable to large systems. Despite the usage of the social instruments, subconventions were still resistant on certain configurations of the environment leading us to perform a more exhaustive study to identify the Self-Reinforcing Structures that are present in social networks like the Scale-Free.

Finally, we have presented a composed social instrument as a robust solution against the persistence of subconventions generated by the identified SRS, improving the convergence times obtained with simple rewiring.

In a world where almost 950 million users belong to an online social networking platform (where virtual agents could also exist) [Radwanick, 2010][9], it is important to understand what mechanisms this virtual entities should be equipped with to facilitate the emergence of common conventions (for the sake of the whole group) as quickly as possible. Moreover, as a system manager, the results from this work highlights the harmful potential of Self-Reinforcing Structures within the network for delaying the emergence process, and draws our attention to solutions for such critical problems.

[9]This report was accessed Sept 1st, 2010 at `http://www.comscore.com/Press_Events/Press_Releases/2010/8/Facebook_Captures_Top_Spot_among_Social_Networking_Sites_in_India`

Chapter 5

Distributed Punishment

The previous chapter has illustrated us how conventions can be delayed by the emergence and maintenance of subconventions. As analyzed, subconventions are mainly generated by the topological configuration of the interaction network. However, this situation emerges as we dealt with the most strict and pure definition of convention, where all the options are potentially equally good as long as the whole population follows the same convention. Consequently it is in the agents' self-interest to accept the convention with which they obtain maximum benefit. However, empirical results made us learnt that certain agents obtain a benefit from following a convention, but due to some exogenous reasons they can be positioned in a frontier.

Being located in a frontier forces an agent to maintain the convention with which the highest benefit is obtained; because of the social learning approach, agents obtain a larger benefit from the convention followed by the majority of their neighbours. The topological configuration can produce the neighbours not to change their convention so our frontier agent interacts successfully with all its neighbours. Because of that reason, some of the interactions of the frontier agent (depending on the number of uncoordinated neighbours) will be unsuccessful. At that moment, all the agents related to the agent(s) in the frontier are interested in reaching a common convention (and resolve the subconvention), to reduce the number of unsuccessful interactions, and maximize their utility.

Agents' strategy are ruled by a utility-maximizing policy. Therefore, by providing agents with the correct mechanisms, they would be able to reduce others utility depending on their behaviour. This type of action is commonly known as a punishment or a sanction[1]. Normally, punishment implies a cost for both the punisher and the punished, reducing both utilities. In that way, we can see how a group of agents could coordinate for punishing a certain agent, whose behaviour should be changed for the benefit of the society, in our case, the frontier agent. Punishment and sanction make special sense in public-good scenarios, where it is in the participants' self-interest to free-ride; however, through punishment/sanctions free-riding can become the least convenient action.

The usage of punishment and sanctions is essential for self-policing so-

[1] We will see in Chapter 6 the difference between punishment and sanction at a cognitive level.

cieties [Papaioannou and Stamoulis, 2005, de Pinninck Bas et al., 2010]. Axelrod [Axelrod, 1986] identified a number of mechanisms to impose social norms in a society: metanorms, dominance, internalization, deterrence, social proof, membership, law and reputation. We are specially interested in dominance and deterrence, in the form of *sanctions*. Posner and Rasmusen [Posner and Rasmusen, 1999] claim that there exist the following types of sanctions:

1. *Automatic Sanctions*: Those that an agent receive for not being coordinated with the others, i.e. those for not following the convention.

2. *Guilt*: The violator feels bad about his violation as a result of his education and upbringing, quite apart from external consequences. Probably most people in our society, though certainly not all, would feel at least somewhat guilty about stealing even if they believed they were certain not to be caught. This can be understood as a self-punishment imposed after breaking an internally respected norm.

3. *Shame*: The violator feels that his action has lowered himself either in his own eyes or in the eyes of other people. In its most common form, shame arises when other people find out about the violation and think badly of the violator. The violator may also feel ashamed, however, even if others do not discover the violation. He can imagine what they would think if they did discover it, a moral sentiment which can operate even if he knows they will never discover it. Also, he may feel lowered in his own eyes, a "multiple self" situation in which the individual is both the actor and the observer of his actions. In order to be able to apply this kind of punishment in MAS, it would be needed to develop into our agent's mind a mental representation of the other agents normative beliefs, being then able to know when others might consider that an agent have broken a norm, and the degree of importance of its violation.

4. *Informational sanctions*. The violator's action conveys information about himself that he would rather others not know. This type of punishment is closely related to trust and reputation systems in MAS. By transmitting a certain piece of information, agents can construct an specific opinion of other agents, to be used during the decision making when interacting with that agent.

5. *Bilateral costly sanctions*. The violator is punished by the actions of and at the expense of just one other person, whose identity is specified by the norm. The expense to that person could be the effort needed to cause the violator disutility, or the utility that the person imposing the punishment loses by punishing him. Examples of what we are calling bilateral costly sanctions could be the agent that spends part of its bandwidth to reduce other agent's bandwidth after sharing a bad quality resource in a P2P network.

6. *Multilateral costly sanctions*. The violator is punished by the actions and at the expense of many other people. One good example would be a coalition of agents reducing another agent's bandwidth for sharing bad quality resources in a P2P network.

Due to the type of assumptions we make in this thesis about the scenario and for which we are designing our utility-based agents, both costly sanctions (bilateral and multilateral) are interesting mechanisms as they can reduce directly other agents' utility. The rest of sanctions (guilt, shame and informational sanctions) are very linked to the internal beliefs and construction of the agent, and can also be delayed in time, being the sanction hard to interpret (like ostracism resulting from the distribution of a bad reputation).

Even though applicable for conventions, costly punishment makes special sense when dealing with essential norms, which are generally named social norms. Essential norms help agents in fixing the focal action for coordinating in the most beneficial situation for the society (full cooperation). However, agents are tempted by the Nash Equilibrium which brings agents towards defection (as an agent obtains a higher benefit from defecting when the rest are cooperating). However, *"If holding a norm is assumption of the right to partially control a focal action and recognition of other norm holders' similar right, then a sanction is the exercise of that right. A sanction may be negative, directed at inhibiting a focal action which is proscribed by a norm, or positive, directed at inducing a focal action which is prescribed by a norm."* [Coleman, 1998].

Theoretical, empirical and ethnographic studies about costly punishment in human societies have demonstrated that this behavior promotes and sustains cooperation in large groups of unrelated individuals and more generally plays a crucial role in the maintenance of social order [Fehr and Gachter, 2002, Ostrom et al., 1992, Boyd and Richerson, 1992]. However, when talking about peer punishment, experimental results [Dreber et al., 2008] have proven that costly punishment increases the amount of cooperation but not the average payoff of the group, specially, reducing drastically the payoff of those subjects applying the costly punishment. We agree that punishment is therefore a second order public good, as it is executed at someone's costs. These costs affect also to the agents' decision making, and might not be convenient for the potential punisher to punish.

We hypothesize that by dividing the costs of punishment amongst the members of the society, this second-order public good becomes less costly, and therefore, more attractive for the members of the society: at a small costs, the potential benefits are increased as the number of freeriders decrease.

Within the scope of the MacNorms project [MacNorms, 2008], and in conjunction with the *Instituto de Análisis Económico*, we have proven the viability of distributed costly punishment through experimentation with human subjects. The main objective of the experiments is to study the validity of distributed punishment amongst human subjects, in order to latter introduce these technologies in heterogeneous societies populated with human and virtual entities.

To perform the experiments with the human subjects we used HIHEREI: a platform to perform experimental economics experiments with human subjects. As this platform is built on top of an *Electronic Institution*, we can easily provide it with the capability of performing experiments with virtual agents.

5.1 HIHEREI: Humans playing in an Electronic Institution

Electronic institutions are regulated frameworks where agents can participate in in order to achieve a tasks. The actions and interactions performed by agents inside are an electronic institution are prefixed by the system designer, being impossible for agents to perform the non-authorized actions. The set of tools already provided to specify, develop and run electronic institutions do not take into account a way to incorporate humans to participate remotely either alone or together with autonomous agents in the electronic institution. Given that, we have extended the current technological framework to incorporate it. This work is the implementation of the design presented in [Sabater-Mir et al., 2007].

The electronic institutions (eI) paradigm is specially useful for us to solve this problem for several reasons. The first one is that the experimental subjects behaviour can be controlled by eI, allowing them only to perform certain tasks in certain moments (with this functionality we can implement the desired experimental scenario). Moreover, the HIHEREI technology allows us to create a remote connection for human subjects to be (invisibly) represented by a virtual agent in the eI; this remote connection has been created to be accessed through any web navigator, allowing the human subjects to participate in the experiment from anywhere in the world[2]. Finally, HIHEREI and the eI provide us with the correct environment for humans to participate with other entities without knowing the nature of them (virtual or human), transporting our subjects of study to the situation that we are more interested: those scenarios where the human subject is not certain of the nature of his peers.

Figure 5.1 shows the main elements of the presented architecture:

- **Electronic Institution (eI)**: The concept of *electronic institution* ([Noriega, 1997, Rodríguez-Aguilar, 2003, Esteva, 2003]) is inspired in human institutions. In open multi-agent systems you have also autonomous entities that interact to achieve individual goals. The behaviour of these entities cannot be guaranteed. Therefore, and similarly to what happens in human societies, agents need mechanisms to guarantee the good functioning of the system in spite of the local behaviours. To fully understand the electronic institution machinery we refer the reader to [Sierra et al., 2004].

- **Virtual agents (E-Agents)**: Agents endowed with autonomous behavior that participate in the e-institution through the *governors* (elements that provide the agent with the interface to interact with the eI and at the same time restrict the possible actions the agent can perform given the current state of the eI).

- **Interface agents (I-Agents)**: Agents that represent human users in the electronic institution. They also act as web servers and, like E-Agents, connect to the eI by interacting with the governors. This allows a totally distributed approach.

[2]In order to test how populations from different countries behave with respect to the punishment we initially planned to run the experiments in the LINEEX Laboratory in Valencia and in the Nutfield Centre for Experimental Social Sciences in Oxford. In the end, because of the shortage of money and the discoveries from other researchers [Noussair et al., 2003], the experiments were restricted to the LINEEX.

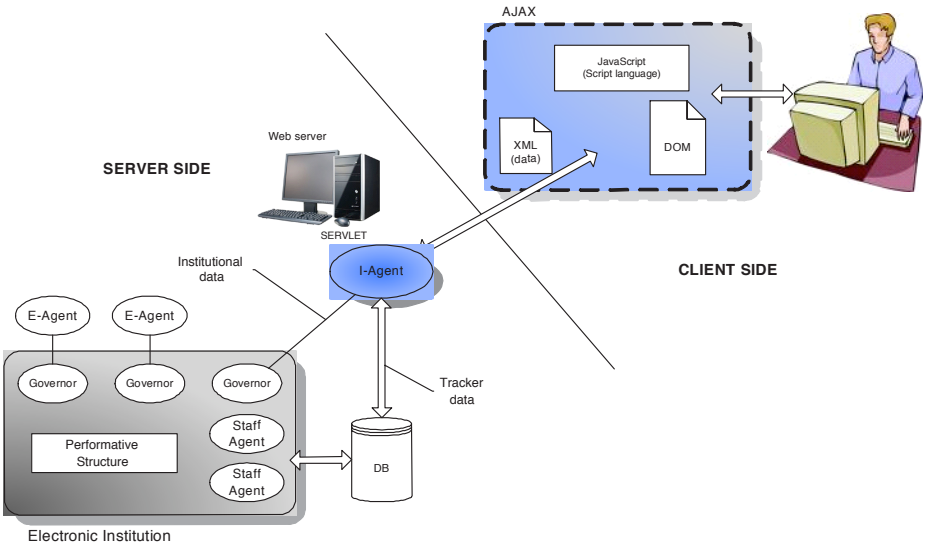


Figure 5.1: Human - eI interaction

- **Staff agents**: Institutional agents in charge of different aspects related to the well functioning of the eI.
- **Data base**: Everything relevant that happens in the e-institution is stored in the DB for a subsequent analysis. The DB also stores the human user activity regarding the actions she/he performs in the client side.
- **Client application**: It is the software that provides a friendly interface for the human user.

5.1.1 Extending EIDE

EIDE [Esteva et al., 2008], the Integrated Development Environment for Electronic Institutions, is a set of tools developed at the IIIA-CSIC aimed at supporting the engineering of multi-agent systems as electronic institutions. However, these tools do not provide a general framework for a flexible participation of humans in electronic institutions. The desirable requirements for this extension are the following:

1. Allow remote access of human users to the eI.
2. Provide flexible and standard client applications for this access. Web applications are a good solution since they can run with standard Internet browsers.

The central element in the link between the human user and the e-institution is the servlet, which is activated once the user (client side), using a web browser, establishes the connection with the server side. This piece of software (I-Agent) is seen as a servlet

from the point of view of the web server but at the same time as a normal agent from the point of view of the eI.

The I-Agent, acting as a servlet, receives messages in XML format from the client application running locally in the user's computer as a web application. The XML messages can be of two types:

- Tracker messages. This is the information that can be used later to analyze the actions performed by the user in the client side.
- Institutional messages. Information that is associated with the e-institution. These are actions that the user wants to perform and that have an influence in the state of the eI.

Simultaneously, the client application receives XML messages from the servlet describing the changes that have been produced in the eI. These changes are shown to the user by the client application.

Under the circumstances of our Experimental Economics Experiments, the following instantiation of the platform was developed:

1. **Electronic institutions**: Using the eI development environment [Esteva et al., 2008] an appropriate electronic institution was defined and verified. This includes the specification of interaction protocols, illocution schemas and a concrete ontology.
2. **Client application**: Using HIHEREI we developed a web-based application to perform the the experiment. Then, human users can be remotely connected to the eI through a simple web browser. The application guides human subjects through the experiment while they interact with other agents (human or simulated).
3. **E-Agents**: Virtual agents are defined to achieve certain experimental situations.

5.2 Experimental Design

In order to test the potential of Distributed Punishment, and following the tradition of Experimental Economics, we have designed the following experiment. Participants of the experiment have to play repeatedly a three phase game. All the participants are divided into groups of 4 that remain constant along the 40 rounds of the experiment.

	$0C$	$1C$	$2C's$	$3C's$
C	5	10	15	20
D	10	15	20	25

Table 5.1: Payoff Matrix for the Distributed Punishment Experiment.

Every round is structured in 3 stages with two decisions to be taken by all the participants (humans or virtual): in the *first stage* participants have to decide whether or not to cooperate to a public good, following the payoffs matrix shown in Table 5.1; the

action C means cooperation and the action D means defection. By observing the payoffs described in Table 5.1 we can easily observe how is in the participants self-interest to defect (by selecting the action D).

During the *second stage* all participants are informed of the decisions of the other participants in their group, and the benefits obtained so far by each of the participants; moreover, subjects are given the option to assign punishments to the other participants within their group, reducing to zero their payoffs obtained in the first stage.

Depending on the treatment, punishing have different costs for the punishers, but it affects in the same way to the punished: removing all the benefits obtained in the first stage of the game. After all the decisions of the second stage have been taken, during the *third stage* all participants are informed of the decisions of the other participants in their group and the resulting payoffs of that round.

In order to test the effect of distributed punishment on experimental subjects we have designed an experiment with 3 different treatments:

1. *No Punishment Treatment*: No punishment is allowed in this treatment. This is the control treatment.

2. *Monolateral Punishment*: In this treatment participants are allowed to send punishments to the other participants assuming each punisher the whole cost of punishment.

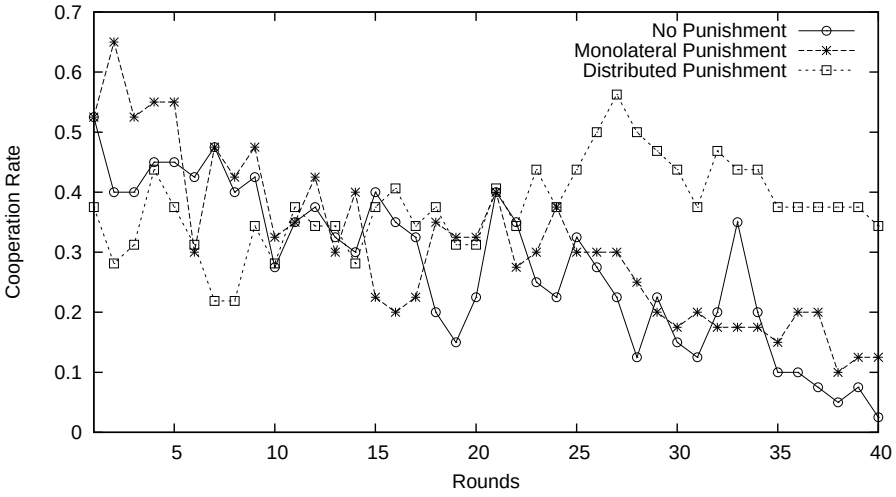
3. *Distributed Punishment*: In this treatment participants are allowed to send punishments to the other participants dividing the punishment costs amongst all the punishers.

In the last two treatments, it is straightforward to observe that the optimal strategy for an individual would be defecting and not punishing. However, this behaviour can be interpreted as freeriding from the rest of peers, imposing costly punishment to the defector. Therefore, with the existence of punishment, subjects can impose it in order to achieve and maintain high cooperation rates. We hypothesize that by having a cheaper option to punish, this punishment will be exerted more frequently, increasing the cooperation rates obtained by the subjects.

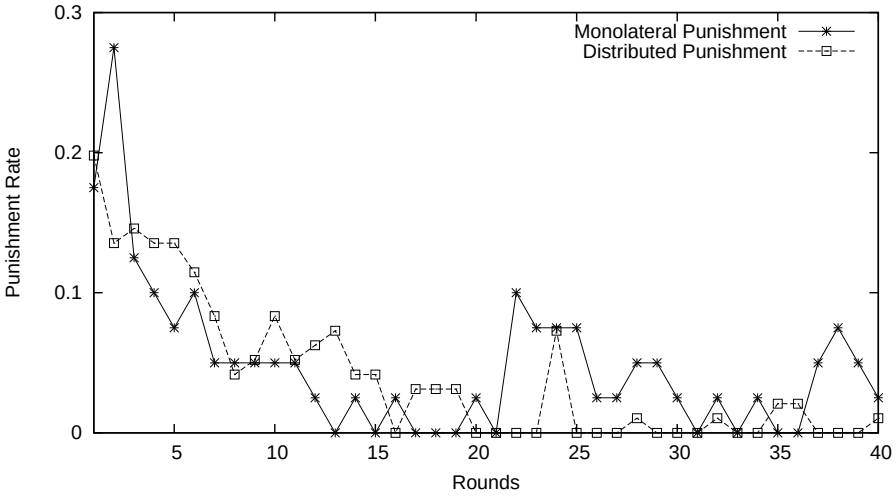
5.3 Empirical Results

The experiments with human subjects were performed in the Laboratory for Research in Experimental Economics (LINEEX), at the Center for Research in Social and Economic Behavior, in the University of Valencia. 40 subjects participated in each treatment, obtaining 10 groups and independent observations per treatment. All the participants were paid depending on their performance on the game.

Experimental results for the 3 treatments are shown in Figure 5.2. Figure 5.2(a) shows the dynamics of the average cooperation rates of the participants in the three treatments, meaning (from [0-1]) the rate of agents that decided to cooperation from the population. Even though high cooperation rates were not obtained in any of the



(a) Average Cooperation Rates



(b) Average Punishment Rates

Figure 5.2: Distributed Punishment Experimental Results with Human Subjects

treatments, experimental results on cooperation rates confirmed that *Distributed Punishment* outperformed *Monolateral Punishment* in a 20,95%, and *No Punishment* in a 31,84%.

We realize that high cooperation rates are not obtained because punishment is still too costly for the subjects, and moreover, the cost of punishment is not linear (in either monolateral or distributed punishment, the relationship between the punishment costs and its effects is not linear, as the punished looses what he earned at the first phase of the game). Other works in the literature have proven that the 1:4 punishment factor is most effective to promote cooperation [Nikiforakis and Normann, 2008]. Therefore we can observe how on the one hand, the punished is affected by the punishment received (as it loses all the benefits obtained in that round), but on the other hand, punishment is too costly to be applied, unless distributed punishment is achieved. Because of the design of the game, distributed punishment is not easily achievable without the possibility of communication amongst the potential punishers, being the punishment rates very low in general, as it can be seen in Figure 5.2(b).

Treatment	*Cooperative Groups*
No Punishment	0
Monolateral Punishment	1
Distributed Punishment	4

Table 5.2: Cooperative Groups in the Different Treatments.

By observing the data more carefully (shown in Table 5.2) we detected in the Distributed Punishment treatment that cooperation was established and maintained in 4 groups, which happen to be exactly the same groups where Distributed Punishment was achieved (in the rest of the groups, subjects did not coordinate for distributed punishment and did not achieve high cooperation rates); we can therefore infer that cooperation was achieved because of the Distributed Punishment. This result is interesting as in the Monolateral treatment, stable cooperation was only obtained in one group, and never obtained in the No punishment treatment.

We would like to remember the reader that the experiment was designed to be executed with no possible communication or planning from the punishers. Agents have to coordinate by acting and only when this happen, punishments costs are reduced for the punishers and therefore cooperation is achieved.

The results obtained regarding the Distributed Punishment were encouraging, however, we noticed that two dynamics were acting at the same time: the punisher's motivations and the punished's motivation. These two dynamics reinforce one to the other, not allowing us to fully understand the behavior of the phenomena.

As we are interested in observing how effective is distributed punishment, we separate both dynamics and analyze carefully the motivations of the punished agent. In order to fully understand the dynamics of the punished subjects, we need to explore the different possible situations, i.e. (1) when the subject receives no punishment, (2) when the subject receives punishment from only one subject, (3) when the subject receives punishment from two subjects, (4) when the subject receives punishment from three subjects.

To extract a scientific conclusion, each situation have to be reached a significant number of times, obtaining the necessary observations; however, it seems unfeasible (in terms of money and time) to repeat the previous experiment as many times as necessary until obtaining the desired results. Moreover, we cannot ensure that the experimental subjects will reach the situations we are interested in by themselves. Therefore, we need to introduce confederate subjects into our experiment, as done in other experiments [Asch, 1955] to produce the experimental situations we need. As the groups in our Public Good Game are of size 4, three confederates are needed for each experimental subject, in order to have a completely controlled situation and observe the subjects' reactions.

For the sake of resource optimization (in terms of economic and time expenses) we decided these confederates to be virtual agents, preprogrammed by us to achieve the different treatments. In that way we fully exploit the capacity of HIHEREI, where the experimental subjects interact in a web-based platform, without knowing the identity of their opponents. Moreover, these type of experiments place the subjects in the type of situation that we are more interested: human subjects playing against other participants without knowing if they are human or virtual entities, as it happens in any interaction normally performed on the Internet.

5.4 Disentangling Distributed Punishment: the Punished's motivation

A total of 80 subjects from the *Universitat Autònoma de Barcelona* participated voluntarily in a repeated Public Good's game at the Experimental Economics Lab. The participants, students from a wide range of fields of study, interacted anonymously using the HIHEREI software. Subjects were not allowed to participate in more than one session of the experiment. In all, four sessions were conducted in December 2010, with an average of 20 participants.

Each treatment was formed by 20 human subjects. Reproducing the conditions of the previous experiment, each human subject was assigned in a group with three other preprogrammed confederate agents which remained the same for the entire session. Subjects knew that the experiment would last 40 rounds, but did not know neither the existence of virtual entities nor the identity of the other participants in their group. The structure of the round follows the same structure of the initial experiment: a first stage were subjects have to decide whether to cooperate or not, a second stage were they decide whom to punish, and a third stage were subjects are informed with the decisions from their peers.

In order to produce a complete experiment, the search space has to be exhaustively explored. To do so we programmed two types of acting strategies for the virtual agents interacting with each subject: the punisher and the non-punisher. Their behavior is programmed to be the same during the first ten rounds, acting differently after the tenth round. The first ten rounds, all virtual actors contribute 50% of the times and punish 25% of the other agents. After the tenth round, all agents contribute 90%. The difference comes in the punishment strategy: non-punishers never punish and punishers

act on defectors 90% of the times only when the punisher has cooperated (to maintain consistency). In order to have complete information about the different possibilities, we explore the complete search space with 4 different treatments with respect the studied human subject:

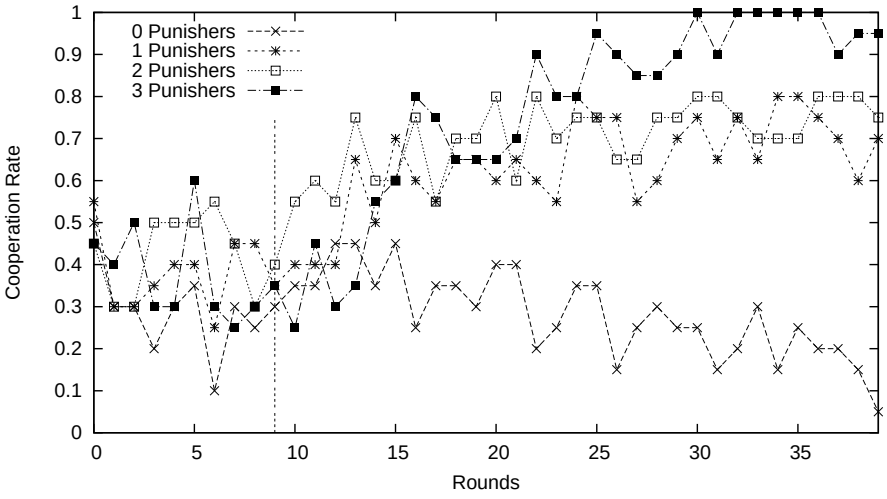
1. *No Punisher Agent Treatment*: each human subject interacting with 3 non-punisher agents.
2. *1 Punisher Agent Treatment*: each human subject interacting with 1 punisher agent and 2 non-punisher agents.
3. *2 Punisher Agents Treatment*: each human subject interacting with 2 punisher agents and 1 non-punisher agent.
4. *3 Punisher Agents Treatment*: each human subject interacting with 3 punisher agents.

Empirical data (shown in Figure 5.3(a)) confirms that distributed punishment have a much stronger impact on the cooperation rates of human subjects, specially when the punishment is consistent from the whole group (3 Punisher Agents Treatment). A small difference between the subjects is taken into consideration when analyzing the results: as the number of punisher actors increase, so does the probability of a subject to be punished (as the punisher actor punishes 90% of the times); the probability of a defector of being punished in each of the treatments is therefore 90%, 99% and 99,9% respectively. These probabilities affect directly to the ratio of non-punished defectors, shown in Figure 5.3(b), along the experiment, that remains below a 5% in the *1 punisher* treatment, around 2% in the *2 punishers* treatment, and a 0,005% in the *3 punishers* treatment. Even though the different punishment probabilities in each treatment, we observe that they do not have significant impact on the non-punished defectors rates to explain the differences in the cooperation rates.

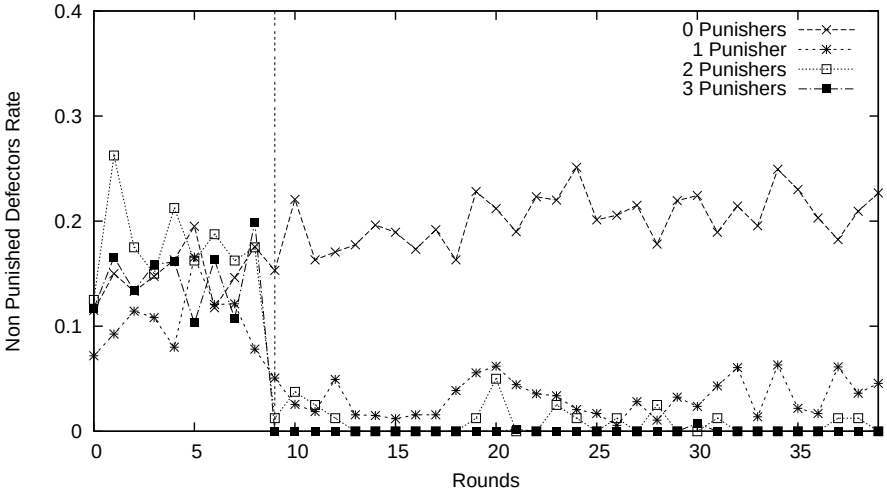
We hypothesize that the explanation for the difference on the cooperation rates is found on the non-evident causes that made the experimental subjects change their cooperation strategy. Punishment affects equally to the experimental subjects at a monetary level independently of the amount of punishers (for those treatments with punishers); however, the subjects decisions have had to be taken by considering other information than the potential monetary punishment. In the treatment with three punishers, cooperation rates are higher than any other treatment.

Based on this empirical result, and those obtained by others [Noussair and Tucker, 2005a], we hypothesize that some behaviours (like a consistent punishment from the entire group) might contain implicit normative messages that affect to the decision making in a more profound way than a mere benefit-cost calculation. *Human subjects do not only take into consideration the potential punishment that can be received, but they rather interpret some social cues as the existence of a social norm, that even in the absence of punishment systems, can lead the subject to abide by the norms*.

As far as we are concerned, there is no existing agent architecture that is able to identify these type of cues and uses them into its decision making. Consequently, we have envisioned a cognitive architecture that incorporate a more complex reasoning than



(a) Average Cooperation Rates



(b) Non Punished Defectors Rates

Figure 5.3: Empirical Results of the Punished.

a simple benefit-cost calculation. This architecture, named EMIL-I-A, will be presented in the following chapter.

Chapter 6

EMIL-I-A: The Cogno-Normative Agent Architecture

In centralized systems norms are explicitly specified ([Garcia-Camino A, 2006, Aldewereld et al., 2007]), might be accessible in a repository and are imposed by a legitimated authority. Differently, in self-organized societies, norms are created and controlled (through peer punishment) by the members of the society. As we discussed in Chapter 1, in this work we focus on the establishment of social norms and not in their emergence.

Peer punishment [Casari, 2004] is considered as a second order public good, as through the application of it, norm compliance is achieved, although it implies a cost for the punisher [Ostrom et al., 1992, Fehr and Gachter, 2000]. This view of punishment is in line with the one supposed by the economic model of crime (see [Becker, 1968]) and with the approach adopted by experimental economics (see, [Sigmund, 2007], for a review of this approach). If punishment is therefore seen only as a deterrent mechanism, in our previous experiment, the three treatments should have provided similar results in terms of cooperation. However, empirical data showed that one of the treatments outperformed the rest, even though the punishment affected in the same way to the instrumental reasoning of the punished.

As noted by others [Giardini et al., 2010, Xiao and Houser, 2005, Hirschman, 1984], peer punishment, and punishment in general, can be seen as a twofold mechanism: (1) as an economic instrument to achieve deterrence on the wrong-doer, and (2) as a signalling mechanism used to let the audience know that the punished behavior is understood as an infraction and therefore it should not be done.

In the experiments performed in the previous chapter, we have seen how the deterrent effect on the wrong-doer is the same in all the treatments, therefore we hypothesize that the difference on the compliance rates can only be explained through a more complex cognitive mechanism.

Nonetheless, several experiments show that punishment is effective in

promoting cooperation also when it does not impose an incentive and it is purely symbolic [Sunstein, 1996, Tyran and Feld, 2006, Houser and Xiao, 2010, Galbiati and D'Antoni, 2007, Masclet et al., 2003]. Social disapproval, peer pressure, public embarassement of offenders, or the communication that a norm violation occurred are often applied to alter people's conduct and in many cases appear to be effective. In these cases, punishing (explicitly or implicitly) informs the offenders and public that the targeted behavior is not condoned, and thus elicits the social norms.

These evidences seem at odds with the idea that people's decisions are influenced only by incentives and they suggest that there are other factors driving their choices. As shown by several works in psychology, *focusing* people's attention on the norm is a crucial factor in producing norm-compliant behavior [Cialdini et al., 1990, Bicchieri, 2006]. One possible explanation of this phenomenon is that the normative content expressed by punishment has the effect of activating people's normative motivation to comply with the norm of cooperation. When the normative content is signalled by punishment, it has the effect of framing the situation in such a way that not only preferences to avoid costs are activated, but normative preferences as well. With normative motivation, we refer to the fact that people follow the norm because they recognize that there is a norm and they are disposed to comply with it even when there is little possibility of instrumental gain, future reciprocation, and when the surveillance rate is very small. We are not claiming here that people are always and only driven by normative motivations when complying with norms, but that their decision is motivated by a combination of normative and cost-avoidance motivations. The hypothesis that people can follow norms as ultimate ends is controversial (see [Barkow et al., 1995], for an evolutionary explanation), but there are several interesting models, such as for example [Bicchieri, 2006, Gintis, 2003], showing how normative preferences can be included in the utility function of the individuals and how this preference interacts with other preferences of the individual.

Recently, researchers have conducted several experiments designed to explore the norm-signalling effect of sanction in the achievement of cooperation, analysing what factors might impact the expressive power of this mechanism [Xiao and Houser, 2005, Masclet, 2003, Noussair and Tucker, 2005a], including ourselves inside the MacNorms project. However, and up to our knowledge, this is a first attempt to fill a gap in the state-of-the-art normative agent literature by designing an agent architecture able to interpret certain social cues (like punishment) in the similar way as humans do. With this contribution we are producing an agent that is able to interact consistently in virtual hybrid environments (where both humans and virtual entities interact).

Basing our work on the findings of cognitive scientists [Giardini et al., 2010], and the results obtained within the MacNorms project, we have developed a BDI cognitive architecture that allow agents to process (and include into its reasoning and decision making) the intrinsic signalling with which punishment might be accompanied, disentangling punishment and sanction.

6.1 Punishment and Sanctions

As specified by [Giardini et al., 2010], punishment and sanction are both mechanisms aimed at changing the behaviors of others, in order to make them abstain from future violations. Because of this similarity, these two phenomena are often mistaken one for another and considered as a *single* behavior. Punishment and sanction are different behaviours and that can be distinguished on the basis of their mental antecedents and of the way in which they aim to influence the future conduct of others.

On the one hand, punishment is referred to be a practice consisting in imposing a cost on the wrongdoer, with the aim of deterring him from future offenses. Deterrence is achieved by modifying the relative costs and benefits of the situation, so that wrongdoing becomes a less attractive option. The effect of punishment is achieved by influencing the *instrumental* mind of the individual, by shaping his material payoffs. This approach to punishment is in line with the economic model of crime, also known as the rational choice theory of crime [Becker, 1968], claiming that the deterrent effect of punishment is caused by increasing individuals' expectations about the price of non-compliance. A rational comparison of the expected benefits and costs guides criminal behaviors and this produces a disincentive to engage in criminal activities.

On the other hand, sanction works by imposing a cost, as punishment does, and in addition by *communicating* to the target (and possibly to the audience) both the existence and the violation of a norm [Giardini et al., 2010, Hirschman, 1984, Xiao and Houser, 2005, Andrighetto et al., 2010b])[1] and at *asking* them to comply with it in the future.

The sanctioner ideally wants the sanctioned to change his conduct not just to avoid the penalty but because he recognizes that there is a norm and wants to respect it. Sanction mixes together material and symbolic aspects and it is aimed at changing the future behaviour of an individual by influencing both its *instrumental* and *normative* mind. In order to decide how to behave, the individual will take into consideration not only a mere costs and benefits measure but also the norm.

Often the sanctioner uses scolding to reign in free-riders, or expresses indignation or blame, or simply mentions that the targeted behaviour violated a norm. Through these actions, he aims to focus people's attention on different normative aspects, such as: (a) the existence and violation of a norm; (b) the high rate of norm surveillance in the social group; (c) the causal link between violation and sanction: "you are being sanctioned because you violated that norm"; (d) the fact that the sanctioner is a norm defender. As suggested by works in psychology, all these normative messages have a key effect in producing norm compliance and favouring social control as well. Even a strong personal commitment to a norm does not predict behaviour if that norm is not activated or made the focus of attention [Bicchieri, 2006, Cialdini et al., 1990]. Furthermore, the more these norms are made *salient*, the more they will elicit a normative conduct.

Norm salience indicates to an individual how operative and relevant a norm is within a group and a given context [Andrighetto et al., 2010b]. It is a complex function, depending on several social and individual factors. On the one hand, the actions of others

[1] Clearly, also punishment can have a norm-signalling effect as an unintended by-product, but only the sanctioner intentionally has this norm-defense goal.

provide information about how important a norm is within that social group. On the other hand, norm salience is also affected by the individual sphere, it depends on the degree of entrenchment with beliefs, goals, values and previously internalized norms of the agent.

We claim that both punishment and sanction favor the increment of cooperation in social systems, but sanction achieves cooperation in a more stable way and at a lower cost for the system. Cooperation is expected to be more robust if agents' decisions are driven not only by instrumental considerations but are also based on normative ones. Moreover, an individual that complies with the norm for internal reasons is also more willing to exercise a special form of social control as well, reproaching transgressors and reminding would-be violators that they are doing something wrong.

6.2 The cognitive dynamics of norms

Building on Ullman-Margalit's definition of a norm [Ullman-Margalit, 1977] as *a prescribed guide for conduct which is generally complied with by the members of society*, a norm has been defined [Andrighetto et al., 2007, Campennì et al., 2009] as a behavior that spreads through a given society to the extent that the corresponding prescription spreads as well, giving rise to a shared set of *normative beliefs* and *goals*. A normative belief is a mental representation, held to be true in the world, that a given action is either obligatory, forbidden or permitted for a given set of individuals in a given context. On the other hand, a normative goal is an internal goal[2] *relativized* [3] to a normative belief: *it is the will to perform an action because and to the extent that this is believed to be prescribed by a norm*.

There are at least three main types of *normative beliefs*:

- the main normative belief, stating that: *there is a norm prohibiting, prescribing, permitting that...* [von Wright, 1963, Kelsen, 1979, Conte and Castelfranchi, 2006, Conte and Castelfranchi, 1999].
- the normative belief of pertinence, indicating the set of agents on which the norm is impinging.
- the norm enforcement belief, indicating that a positive sanction is consequent to norm obedience and a negative sanction is consequent to norm violation.

In order to be compliant with the norm, the first two normative beliefs are necessary conditions: agents should recognize that there is a norm and that it applies to them. When individuals do not have them in mind, norms exert

[2]From a cognitive point of view, goals are internal representations triggering-and-guiding action at once: they represent the state of the world that agents want to reach by means of action and that they monitor while executing the action [Conte, 2009]

[3]A goal is relativized when it is held because and to the extent that a given world-state or event is held to be true or is expected [Cohen and Levesque, 1990]. An example is the following: tomorrow, I want to go sunbathing to the beach (relativized goal) because and to the extent that I believe tomorrow it will be sunny (expected event). The precise instant I cease to believe that tomorrow it will be sunny, I will drop any will to go to the beach

no effect on the behavior [Andrighetto et al., 2007, Campennì et al., 2009]. Furthermore, the more these normative mental representation are salient, the more they will elicit a normative behavior [Bicchieri, 2006, Xiao and Houser, 2005, Cialdini and Goldstein, 2004]. Norm compliance and norm salience are strongly intertwined: findings from psychology [Cialdini et al., 1990, Bandura, 1991] and behavioral economics [Xiao and Hauser, 2009, Bicchieri and Chavez, 2010] have pointed out that drawing people's attention on a social norm and making it salient elicits an appropriate behavior. Making a norm salient typically means providing people with information about the behavior and beliefs of the other individuals [Bicchieri and Xiao, 2007, p 4].

Despite the main normative belief and the belief of pertinence, the norm enforcing belief is not a defining element of the norm, it simply enforces it. A *normative command* is a special command that is intended to be adopted by its addressees because it is normative and *norms must be obeyed* [von Wright, 1963]. Of course this motive can be absent or weak in the minds of people, depending on the socialization and education process and the credit obtained by current institutions. *Sub-ideally*, norms are often complied with because they are enforced by a system of sanctions. But *ideally*, they are meant to be observed because are norms and should be complied with for their own sake.

However, a belief is not yet a decided action. Normative beliefs are necessary but insufficient conditions for norms to be complied with. What leads agents endowed with one or more normative beliefs to execute them, especially since, by definition, norms prescribe costly behaviours? How can norms generate goals?

Usually normative beliefs generate normative goals by reference to an external enforcement[4] (sanctions, approval, etc.). The agent calculates the costs and benefits of complying with or violating the norm and then decides how to behave. If no such a goal is generated, the norm will be violated.

6.3 The EMIL-I-A Architecture

In order to account for the different forms of punishment (and intended to achieve other cognitive processes as we will see) a rich cognitive platform, namely a BDI-type architecture is required and EMIL-A [Andrighetto et al., 2007, Campennì et al., 2009] seemed a good candidate. Our extension is called EMIL-I-A, which stands for *EMIL Internalizer Agent*. As the name suggests this architecture was first conceived for Norm Internalization. Internalization is achieved with another module of this architecture that will be carefully analyzed in Chapter 8. In this chapter we center on how the difference between sanction and punishment is taken into account by for the described architecture.

EMIL-I-A (as EMIL-A) consists of mechanisms and mental representations allowing norms to affect the behavior of autonomous intelligent agents. As any BDI-type (Belief-Desire and Intention) architecture EMIL-I-A operates through modules for different sub-tasks (recognition, adoption, decision making, salience control, etc...) and acts on mental representations for goals and beliefs in a non-rigid sequence.

[4] See [Conte and Castelfranchi, 1995, Conte and Castelfranchi, 2006] for a fine grained analysis of different reasons behind norm compliance)

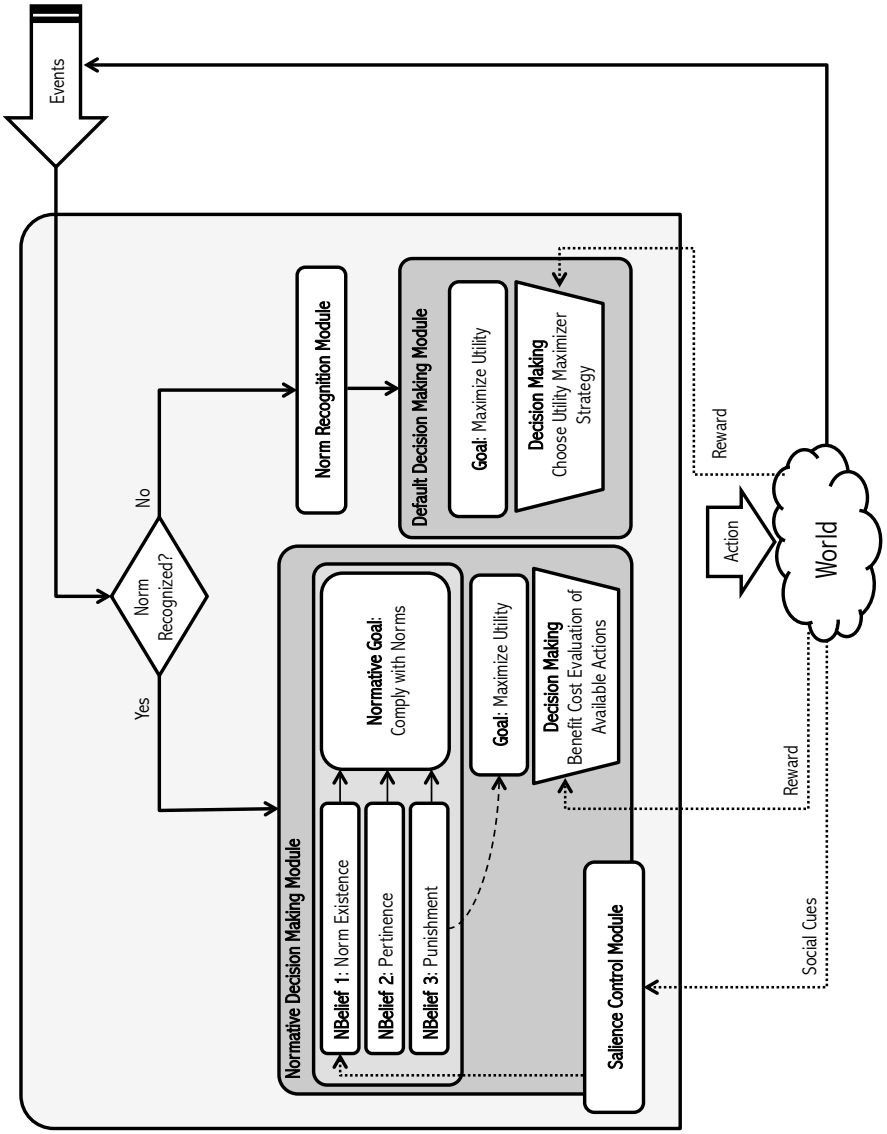


Figure 6.1: EMIL-I-A Architectural Design

Our normative agent architecture (represented in Fig. 6.1) has three important parts: the *norm recognition module*, the *salience control module*, and *decision-making*.

For further references on how the norm recognition module works, we refer the reader to [Campennì et al., 2009]. The *norm recognition module* (that is accessed when norms have not yet been recognized, as it is shown in Fig. 6.1) allows agents to interpret a social input as a norm. In order for agents to recognize the existence of a norm, they have to hear from consistent agents[5] a certain number of normative messages, such as "you should not take advantage of your group members by shirking" and observe a few normative actions compliant with the norm or aimed to defend it (i.e. cooperation, punishment and sanction, observed or received).

After recognition, a norm activates the three types of normative beliefs described in Section 6.2, that are stored in the *normative board*[6] (inside the Normative Decision Making Module represented in Fig. 6.1) . Once generated or activated, normative beliefs will be inputted to the norm-adoption module: a normative goal - relativized to the expected enforcement - will be generated. In this condition the normative agent adopts the norm, because it wants to avoid punishment. The normative goal is then inputted to the normative decision-maker and compared with other goals (on the basis of their salience, updated with the Salience Control Module) possibly active in the system. The normative decision-maker will choose which one to execute and will convert it into a normative intention (i.e. an executable goal). Once executed, this normative intention will give rise to norm-compliance and/or norm-defense and/or norm transmission through communication. Otherwise, it will eventually be abandoned, solution that brings to the utility maximizing strategy (normally, norm violation).

6.3.1 Norm Salience

We adopt a definition of norms in which social norms are not static objects and the degree to which they are operative and active varies from group to group and from one agent to another: we refer to the degree of activation as norm's *salience*. The more salient a norm is, the more it will elicit a normative behaviour [Bicchieri, 2006, Xiao and Houser, 2005, Cialdini et al., 1990]. Norm's salience is a complex function, depending on several social and individual factors allowing agents to *dynamically* monitor if the normative scene is changing and to adapt to it[7]. For example, in an unstable social environment, if norm enforcement suddenly decreases, agents having highly salient norms are less inclined to violate them, as a highly salient norm is a reason for which an agent continues to comply with it even in absence of punishment. It guarantees a sort of *inertia*, making agents less prompt to shift from the present strategy to a more favorable one. Viceversa, if a specific norm decays, agents are able to detect this change, ceasing to comply with it and adapting to the new state of affairs. Finally, if an agent faces an emergency or a normative conflict, norm salience allows it to decide which action to perform providing it with a

[5] An agent is consistent if, when choosing to punish, it has before cooperated in the PD.

[6] The normative board is a portion of the long-term memory where normative beliefs are stored, ordered by salience.

[7] It is interesting to notice that this mechanism allows agents to record the social and normative information, without necessarily proactively exploring the world (e.g. with a trial and error procedure).

criterion to compare the norms applicable to the context. For example, on the basis of the relative salience of the two norms, "stop at the red traffic light" and "do not block the way, when listening to an ambulance siren", the agent will decide whether to wait or move on.

This is possible because our normative agents are as autonomous as socially responsive. They are autonomous in that they act on their own beliefs and goals (on the basis of their salience). However, they are also responsive to their environment, and to the inputs they receive from it, especially to social inputs.

On the one hand, the actions of others provide information about how important a norm is within the group. In particular, norm's salience is affected by:

1. The amount of observed compliance and the costs people are willing to sustain in order to comply ([Cialdini et al., 1990]).
2. The surveillance rate, the frequency and intensity of punishment ([Haley, 2003]).
3. The enforcement typology (private or public, 2nd and 3rd party, punishment or sanction, etc.) ([Masclet, 2003]).
4. The efforts and costs spared to educate the population to a certain norm; the visibility and explicitness of the norm ([Cialdini et al., 1990]).
5. The credibility and legitimacy of the normative source ([Sacks et al., 2009]).

On the other hand, norm salience is also affected by the individual sphere, it depends on the degree of entrenchment with beliefs, goals, values and previously internalized norms. It has to be pointed out that the norm salience can also gradually decrease: for example, this happens when agents realize that norm violations received no punishment or when normative beliefs stay inactive for a certain time lag, suggesting that the norm is not very active in the population anymore.

In order our agents to obtain an accurate and representative salience measure, we have designed a *Salience Control Module*. This module is fed with the social information accessible by agents, transforming and aggregating those cues into a value that represents the degree of activation of an specific norm. With that information, agents can be endowed with other mechanisms in order to decide when to abide by (or violate) intelligently the existing social norms. However, the agents' decision making module is dependent on the scenario and environment and we will analyze two different decision making modules, built on top of norms' salience, in the following chapters.

6.3.2 Salience Control Module

By using the Salience Control Module agents can understand the relative importance of each norm. The module is fed (interaction after interaction) by social and normative cues available within their surrounding, and those are the following:

- *Compliants Observed*: Neighboring agents who *comply* with the norm.
- *Violators Observed*: Neighboring agents who *violated* the norm.

Social Cue	Weight
Norm Compliance/Violation (C)	w_C = (+/-) 0.99
Observed Norm Compliance (O)	w_O = (+) 0.33
Non Punished Violators (NPD)	w_{NPD} = (-) 0.66
Punishment Observed/Applied/Received (P)	w_P = (+) 0.33
Sanctioning Observed/Applied/Received (S)	w_S = (+) 0.99
Explicit Norm Invocation Observed/Received (E)	w_E = (+) 0.99

Table 6.1: Normative Social Information Weights for the Salience Aggregation Function.

- *Non Punished Violators*: The amount of unpunished neighboring agents that violated the norm.
- *Punishment Observed/Applied/Received*: The amount of punishment actions observed in the neighbourhood, applied to other agents or received from other agents.
- *Sanctioning Observed/Applied/Received*: The amount of sanctioning actions observed in the neighbourhood, applied to other agents or received from other agents.
- *Explicit Norm Invocation Observed/Received*: The amount of explicit norm invocations observed in the neighbourhood, applied to other agents or received from other agents.

Each of these cues (see Table 6.1) is aggregated with different weight. A higher weight is given to those social actions that are interpreted as driven by normative motivations, and the values and their ranking have been extracted from [Cialdini et al., 1990].

Norms' salience is updated according to the formula below and the social weights described in Table 6.1:

$$\mathrm{Sal}_t^N = Sal_{t-1}^N + \frac{1}{\alpha \times \phi}(w_C + O \cdot w_O + NPD \cdot w_{NPD} + P \cdot w_P + S \cdot w_S + E \cdot w_E) \quad (6.1)$$

where Sal_t^N represents the salience of the norm N at time t, α the number of neighbors that the agent has, ϕ the normalization value, w_X the weights specified in Table 6.1, and finally O, NPD, P, S, E indicate the registered occurences of each cue. The resulting salience measure ($Sal_t^N \in [0-1]$, 0 representing minimum salience and 1 maximum salience) is subjective thus providing flexibility and adaptability to the system.

6.4 Conclusions

In this chapter we have discussed about the necessity of a cognitive model that is able to represent the existence of norms into the agents decision making. We have seen in previous chapters how an utilitarian-based agent architecture is not enough, based on the empirical results obtained with human subjects.

We have developed a cogno-normative agent architecture based on different normative beliefs and goals that allow agents to comply with norms in a cognitive way when these are recognized. Our modular architecture activates a different decision making when norms are identified (following a normative decision making) different than the classical utility-maximizing one, as those seen in Chapters 3 and Chapter 4. Once the norms are identified, the normative beliefs related to the recognition of a norm will also affect the agents decision making, orchestrated by the norms salience. This salience indicates agents how important norms are within a normative context, allowing agents to perform an intelligent norm violation in those necessary cases.

In the next chapter we will put this architecture into use, exploiting it to observe the effect of norms and punishment into the agents self-regulation.

Chapter 7

Proving EMIL-I-A

In this chapter we explore, by means of cognitive modeling and agent based simulation, the specific ways in which punishment and sanction favor the recognition of social norms, and adapt accordingly their degree of activation, in order to achieve an intelligent compliance.

For such task, we put EMIL-I-A into action. Therefore, we need to develop a complete agent architecture, integrating the *Salience Control Module* (explained in Chapter 6) with the agents' decision making. We prove the efficiency of EMIL-I-A by reproducing with agent based simulation the *Distributed Punishment Experiment* presented in Chapter 5.4.

Based on our results with human subjects in the Distributed Punishment Experiment, and under the suspicion that different types of punishment have a different effect on the emergence of cooperation, we will present the design and results of new experiments with human subjects that make explicit the difference between punishment and sanction, proving afterwards the consistency of EMIL-I-A with the empirical results.

After checking the efficiency of EMIL-I-A we fully exploit its capabilities by analyzing in a simulated scenario the different dynamics of punishment and sanction in the achievement of norm compliance, focusing specially on the emergence of cooperation. As both punishment and sanction are costly behaviors, both for the enforcer and the target, a well-designed enforcement system should combine high efficacy in discouraging cheating with limited costs for society. For such reason we also introduce a heuristic to be used locally by agents in order to intelligently adapt the regulation costs.

7.1 EMIL-I-A plays Distributed Punishment

In order to test the EMIL-I-A implementation we firstly prove its behaviour in the *Distributed Punishment* experiment presented previously.

As it was specified in Section 5.4, in the *Distributed Punishment* game agents have to take two decisions at two different phases on a repeated game. During the first stage agents have to choose whether to cooperate or not, and then in the second stage they have to choose whom to punish inside their group.

The dynamics of the game are exactly the same, and we will not repeat them (referring the reader to Section 5.2 for the complete specification of the experiment). However, we need to specify how the agent take these two decisions.

7.1.1 EMIL-I-A Decision Making

In this model, agents have to take two decisions at two different stages: to cooperate or defect and to punish/sanction or not. These decisions are influenced by an aggregation of economic, social and normative considerations. The decision making modules in charge of taking these decisions are built into the agents following an evolutionary approach using mixed strategies. These mixed strategies are updated after each timestep in the simulation with an aggregation of the *Drives* that are explained in this section. These mixed strategies allow agents to adapt to the environmental conditions, and to other changes in their social environment.

First Stage Decision: Cooperate or Defect?

We have designed three different drives that are aggregated to update the mixed strategy. These three drives represent the influence of different aspects into the cooperation decision.

$$\text{SID}_t = SID_{t-1} + (O \times \frac{R_t - R_{t-1}}{R_{Max} - R_{Min}}) \quad (7.1)$$

1. **Self-Interested Drive (SID)**: it motivates agents to maximize their individual utility independently of what the norm asks. The SID is updated according Equation 7.1.1 where SID_t represents the self-interested drive at time t, O the orientation (+1 if the agent Cooperated and -1 if it Defected), R_t the reward obtained at time t, and R_{Max} and R_{Min} respectively the maximum and minimum reward that can be obtained. In the case where the marginal reward is zero, it is substituted by an inertial value with the same orientation as in the last variation. This way, the proportional and normalized value of the marginal reward (a proportional value of the gain/loss wrt the last timestep) obtained indicates how the agent would change its *cooperation probability*.

2. **Social Drive**[1]: Agents are influenced by what the majority of their neighbors do. Combining the results obtained in the experiments with human subjects, 31% of agents ([Asch, 1955]) will change their tendency (with an increment of 18% [Fowler and Christakis, 2010]) towards that of their neighbors in case those are unanimous.

[1]Even though we model this drive at the theoretical level, we have decided not to include it in the actual platform yet in order to have clearer results.

3. **Normative Drive (ND)**: once the cooperation norm is recognized, agents decisions are influenced also by the normative drive. The normative drive is affected by the norm salience: the more salient the norm is, the higher the motivation to cooperate. Salience is updated as explained in Sec. 6.3.2.

Second Stage Decision: Punish or Sanction or none?

The agent's decision making module is also in charge of deciding when to punish or sanction a non compliant behaviour. Considering the difference between punishment and sanction, agents need to decide which one to use: only agents having recognized the existence of a norm regulating their group can sanction, otherwise they will just use punishment.

As discussed in Section 6.1, the punisher and the sanctioner are driven by different motivations. The former punishes in order to induce the future cooperation of others, thus expecting a future pecuniary benefit from its acts. On the other hand, the sactioner is driven by a normative motivation: it sanctions to favor the generation and spreading of norms within the population. Given these differences, the probability governing the decision of punishing or sanctioning is modified by different factors[2] and they change in the following way:

(4) Punishment Drive: Agents change their tendency to punish on the basis of the relative amount of defectors with respect to the last round. *"The more commonly a norm violation is believed to occur, the lower the individuals' inclination to punish it."* [Traxler and Winter, 2009]. If the number of defectors increased, agents' motivation to punish decreases accordingly.

(5) Sanction Drive: Agents change their tendency to sanction on the basis of the norm salience. The more salient the norm is, the higher the probability to sanction defectors.

Therefore, we can see how the mixed strategies are affected both by agents' decisions and by social information. As in evolutionary game theory, eventually these mixed strategies can tend to extreme values (full cooperation or full defection, and complete punishment or no punishment), thus meaning that the system has converged.

7.1.2 EMIL-I-A Results in Distributed Punishment Experiment

As it was done with the human subjects, each subject (in this case, the EMIL-I-A agents) is assigned with 3 other preprogrammed agents. The effects of the 4 different treatments (no punishment, 1 punisher, 2 punishers, and 3 punishers) are analyzed.

Before showing the results, we need to specify that EMIL-I-As are loaded into the experiment with an initial cooperation and punishment probability of 50% based on empirical evidence. As communication was not allowed in the experiments with human subjects, the EMIL-I-As' *norm recognition module* parameters have been changed, reducing the normative messages to 0 and leaving the other parameter with the same value. Therefore, for an agent to recognize the norm, 10 normative actions need to be

[2]Even though agents pay a cost to enforce defectors, this cost is not taken into account when they update their decisions to punish and sanction.

seen (either compliance, a violation, or a punishment). Moreover, and because of the same previous reason, the confederate actor agents only punish, and not sanction. In this way we are able to reproduce with fidelity the experimental conditions under which human subjects were tested.

Experimental results show that EMIL-I-As (shown in Figure 7.1(b)) behave following the same pattern with respect to humans (whose results are shown in Figure 7.1(a)). However, the difference observed in the experiments with human subjects between the three treatments (with a stronger impact on cooperation when 3 punishers acted simultaneously) is not as strong with EMIL-I-A subjects.

In order to observe whether EMIL-I-A is being affected by the different punishment probabilities noticed in Sec. 5.4, we contrast the results with other type of agents' architectures, driven exclusively by utilitarian motivations (as the ones we used in Chapter 3.1, built with Reinforcement Learning Algorithms).

From the different options (Fictitious Play [Fudenberg and Levine, 1998], Q-Learning [Watkins and Dayan, 1992], Win or Learn Fast - Policy Hill Climbing [Bowling and Veloso, 2002]) we decide to use a WoLF-PHC as it is the one that we believe represents better the learning of an utilitarian-based human: it will maintain its strategy while obtaining benefits, and it will change it and learn otherwise.

Experimental results show that the reinforcement learning agents obtain similar dynamics than humans, confirming that the instrumental motivation in humans is very strong, although the high cooperation rates obtained with humans (and EMIL-I-As) are not reached.

These results give us the certainty that the dynamics of EMIL-I-A have been correctly integrated from the model into a working architecture, allowing us to recover our initial hypotheses and test the specific ways in which punishment and sanction favor the achievement of cooperation and the spreading of social norms in social systems populated by autonomous agents.

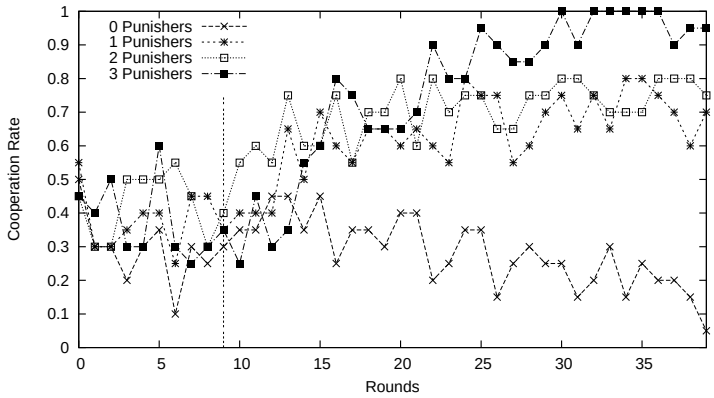
7.2 Experimenting with Punishment and Sanction

Experimental economics experiments with human subjects allowed us to differentiate the effects of punishment and sanction on the emergence of cooperation. As we have commented already, our hypothesis states that when sanction is available as a mechanism for peer punishment, higher cooperation rates are obtained and, in the absence of punishment, remain higher for a longer time than when punishment is used.

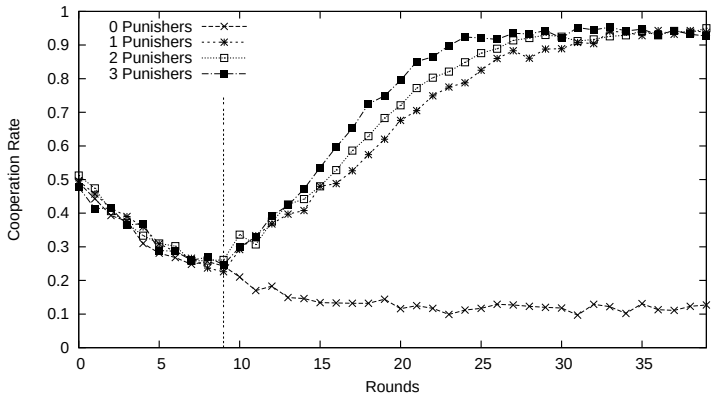
To test this hypothesis we perform a new set of experiments with human subjects that explicitly differentiates between punishment and sanction. For the sake of comparison, the experiment is designed in a similar fashion to [Fehr and Gachter, 2000].

7.2.1 Experimental Design

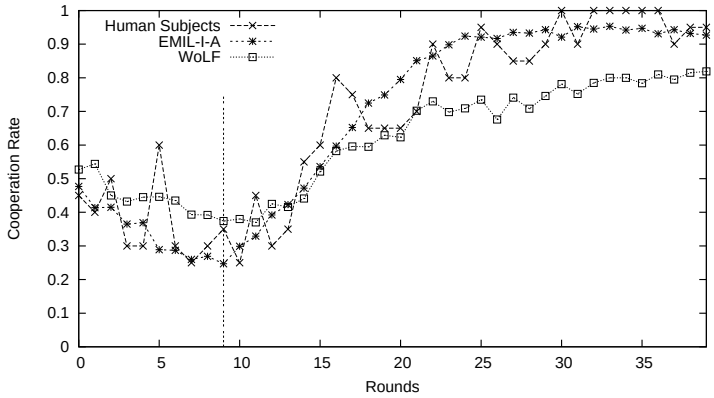
Participants are divided into groups of 4, that remain constant for the whole treatment. For this experiment, no virtual agents were used, and all the experiments were performed following the Experimental Economics methodology [Croson, 2005].



(a) Human Subjects Cooperation Rates



(b) EMIL-I-A Subjects Cooperation Rates



(c) Comparing Humans, EMIL-I-A and WoLF agents

Figure 7.1: Distributed Punishment Experimental Results with Simulated Agents

As the objective of this experiment is to test the difference of punishment and sanction on the subjects decision making, we perform two different treatments: the *Punishment Treatment* and the *Sanction Treatment*. Each treatment is composed of 3 blocks of 10 rounds (with a total of 30 rounds). Depending on the treatment, the second block will vary, remaining the first and third block constant.

In all three blocks, at the beginning of every round, participants are given with an endowment of 20 ECUs, and at the end of each round participants are informed of the benefits obtained in that round.

In the first block of 10 rounds (as well as in the third block), subjects simultaneously choose the portion of their endowment to contribute to a group account. Each ECU contributed to the group account yields a payoff of 0.4 ECU to each of the four members of the group. Each ECU not contributed by the subject is credited to the subjects own earnings. Therefore, the earnings of an individual in a period are calculated with formula 7.2.1, where c_x is the amount contributed by subject x.

$$\text{E} = 20 - \text{c}_i + 0.4 \times \sum_{k=1}^{n} c_k \quad (7.2)$$

Once all participants of the group make their contribution decision for that round, the system informs each subject of his initial endowment, his individual contribution, the total group contribution, and his individual earnings.

In the second block (rounds 11-20), subjects have to take decision in two phases. The first phase decision is exactly the same than in the first block. During the second phase, agents have to simultaneously decide whom and how much to punish/sanction. At the end of the first stage, subjects are informed of the contribution levels of each of the other members of their group. For the *Punishment Treatment*, they can assign from zero to ten punishment points to each of the three other group members, with a cost of 1 ECU per point. Each point received from any other agent reduces the first stage earnings of the receiving subject 3 ECUs.

In the case of the *Sanction Treatment*, the second phase remains equal with a slight variation: apart from sending the punishment points as in the *Punishment Treatment*, subjects can also send a message that justifies their punishment decision and will be sent together with the punishment points. The message is predefined by us, allowing certain parameters to be chosen by the subjects, like the correct amount that should be contributed (represented by the *XXX*). The message structure is the following:

`It should be contributed` *XXX* `ECUs, because:`

1. `it is what is supposed to be done.`
2. `because that way we all are better off.`
3. `because that way you will not be punished.`

With this message structure, subjects have to specify (in the *XXX* provided space) what they consider the *"correct"* amount to be contributed, and one of the three reasons why that should be done. The first message ("*it is what is supposed to be done*") corresponds to a normative explanation, the second message ("*because that way we all are better off*") corresponds to a social preference explanation, and the last one ("*because that way you will not be punished*") corresponds to a self-interested explanation. This way, in the *Sanction Treatment*, subjects have the chance to send punishment points with a justification of what the correct behaviour should be and why, in contrast with the *Punishment Treatment* where no messages are sent. We want to remark that theoretically, and as the messages that go with the punishment points are optional and cost-free, the *Punishment Treatment* can produce the same utilitarian effect than the *Sanctioning Treatment*.

At the end of this second phase, participants are shown the punishments (and messages if any) sent to each of them, and their final benefit of the round.

This three block experimental structure allows us to observe what happens in the absence of punishment (in the first block) and clearly observe the effect of both punishment technologies when available (in the second block). The third block of the experiment allows us to observe the inertia introduced in the subjects with punishment/sanction.

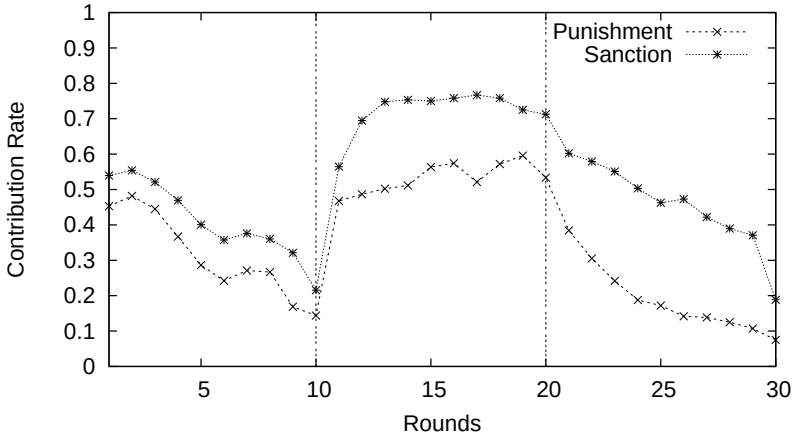
7.2.2 Empirical Results

The experiments with human subjects were performed in the Laboratory for Research in Experimental Economics (LINEEX), at the Center for Research in Social and Economic Behavior, in the University of Valencia. 48 subjects participated in each treatment, obtaining 12 groups and independent observations per treatment. All the participants were paid depending on their performance in the game.

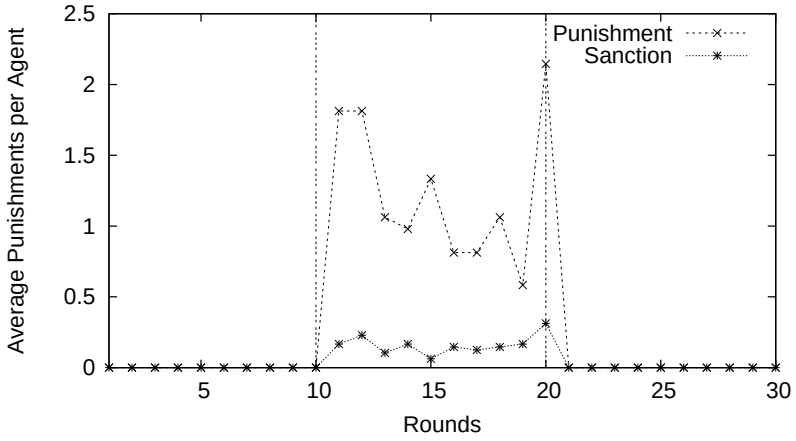
Experimental results are shown in Figure 7.2(a). Empirical data confirm our hypotheses showing that sanction do have a stronger impact on the subjects decision making, resulting in a higher and more stable cooperative strategy than when punishment is used. This result is important since the message accompanying the punishment do not have any monetary costs associated, and indeed produces higher cooperation rates. Moreover, as we can see in Figure 7.2(b), the average amount of sanctions sent per subject is considerably reduced with respect to the average amount of punishments[3].

In general we can observe how both punishment/sanctioning technologies are effective in attaining contribution levels higher than would typically be observed in the absence of any system. The results obtained in our experiments are in concordance with those presented in [Fehr and Gachter, 2000, Noussair and Tucker, 2005b], at least in the punishment treatment. In [Noussair and Tucker, 2005b] authors incorporate a

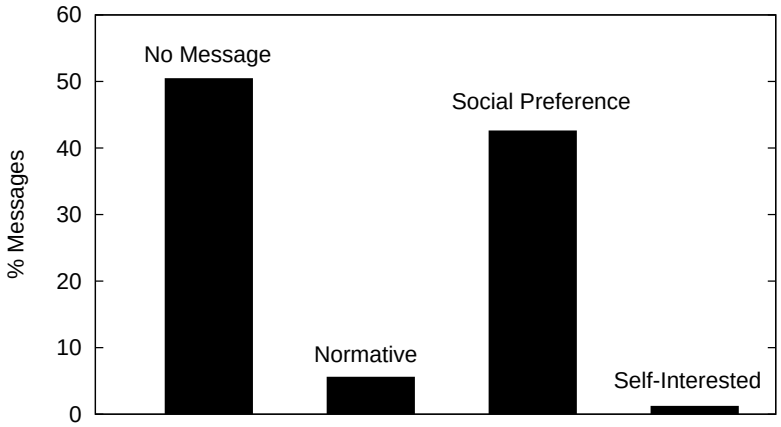
[3]At the beginning of each treatment, subjects are informed about the 3-block structure of the experiment, but they did not know the specific configuration of each block. Instructions were read and explained at the beginning of each block. Because of this experiment configuration, subjects discover only at the beginning of the second block that they can punish/sanction after the first phase, and they will be able to do it for 10 rounds. This introduces an "end-of-the-world" effect, for which, when "the end of the world" is approaching, the decisions are not rational anymore. In our case, it is the last round of the second block, where we observe an abrupt increase of both punishments and sanctions, which should not be considered as representative.



(a) Average Cooperation



(b) Average Punishment/Sanction



(c) Message Distribution for the Sanction Treatment

Figure 7.2: Punishment and Sanction Experiment with Human Subjects

symbolic punishment that reminds of our sanction mechanism: they allow subject to send punishment points in the same way we did in our *Punishment Treatment* although with no costs associated to the subjects assigning or receiving points. These points can be understood as a level of disapproval of a subject's contribution decision in the first stage. However, they lack of an accompanying significance of why that should be done, which is included in the message that goes with our punishment points.

With respect to the message chosen by subjects, in the *Sanctioning Treatment* (the only one where a message can be accompanied with the punishment points), subjects have a preference between sending no message at all (50,48% of the times and obtaining a plain punishment), or, sending the *Social Preference* message (*"because that way we all are better off."* is sent 42,63% of the times), as it is indicated in Figure 7.2(c). While others have studied the strength of different types of messages accompanying punishment with other punishment technologies [Bo and Bo, 2009] (centralized punishment coming from the system instead that from peers), our experimental results confirm that when punishment comes with the message justifying the punishing decision from the peer, cooperation rates are increased. Even though we will not further explore the reasons of this effect (and leaving this line of research open for future work) we hypothesize that this occurs because when punishment is accompanied with a message the punisher is seen as an entitled authority by the punished, activating its normative mind.

With the light thrown with the results obtained with human subjects, we can affirm that our assumption (integrated into the EMIL-I-A platform) about the stronger impact of sanctioning over punishment was correct.

7.2.3 EMIL-I-A Results

In order to check the correctness of the implementation of our EMIL-I-A platform, we decided to simulate the previously performed experiments substituting the human subjects with EMIL-I-A agents, as we did with the *Distributed Punishment* experiment in Sec.7.1.

There are a number of reasons why we believe that it is hard to simulate the human subjects results as there are many other factors affecting their decision making and cannot be controlled in these in-vitro experiments. Among them, the one we consider essential is the distribution of personalities on the experiment; it has already been demonstrated that there are different types of subject attitudes towards contribution in a public good [Brennan et al., 2008], and we believe that this proportion of each type of potential personality has an effect on the group's dynamics.

In the experiments with EMIL-I-As we assume an homogeneous population of agents, with the same characteristics and tendencies in the decision-making; this assumption is taken because for the design of EMIL-I-A, average values from human subjects were taken and implemented in the architecture, without considering different types of human personalities[4].

[4] Agents can be given a "personality" with the EMIL-I-A architecture by fine-tuning the decision making. In the specific configuration detailed in Section 7.1.1, different weights applied while aggregating the different drives would produce a different type of behaviour.

We need to make a number of assumptions that will help us fix some EMIL-I-A's behavioural parameters before running the simulation: agents are initialized with a tendency to cooperate of the 50%, an initial tendency to punish of the 70%. We believe these are pertinent (based on empirical results from humans) and necessary to run the simulation in the proper way. Moreover, we assume that our EMIL-I-A agents already know the existence of the cooperation norm (as humans do before going into the experiment) with an initial salience of 0.75.

In order to reproduce the same experiments performed with human subjects, we restrict the simulation platform to only allow punishment on one of the treatments, and punishment and sanction in the second treatment. For the sanctioning treatment we have taken an implementation decision: as we saw with the human subjects, the message chosen by them was fundamental to achieve the desired effect (cooperation rate's raise); in our architecture, and assuming the homogeneity of agents, we have decided to avoid the problems associated to the semantics of the message and we fix this message as a universal message, to be understood equally by all EMIL-I-A agents as the normative message that is part of a sanction [5].

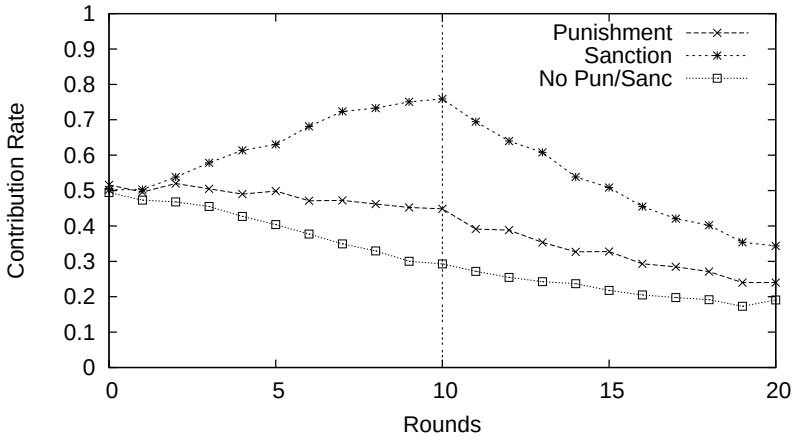
We have slightly altered the structure of the experiment by eliminating the first block: restricting agents from the punishment/sanctioning technologies, cooperation rates will start decreasing. Because of the evolutionary strategies with which agents are provided, the recovery from the absence of punishment/sanction would result to be much slower than in humans. The reason why humans react in a such a way is because the "re-start" effect: the experiment pauses and the instructions are read to humans, producing in them the image of a re-start of the experiment. Unfortunately these dynamics cannot be realistically incorporated into the agents.

The experimental results are shown in Figure 7.3(a). The EMIL-I-A's obtained dynamics when contrasting punishment and sanction are similar than those obtained with humans: when sanctioning is available, cooperation rates reach values between 70% and 80%, in contrast with the 50% obtained with punishment. Moreover, in the absence of punishment, cooperation rates in both cases start decreasing, being the decreasement slower when sanction is used. In general, we observe that EMIL-I-A is slower than humans because of the evolutionary strategy update performed; on the other hand, humans were affected by different external inputs like the instructions reading between blocks, and possibly other non-identified dynamics. The delay observed is of 5 rounds, being the average behaviour almost identical after that adaptation time.

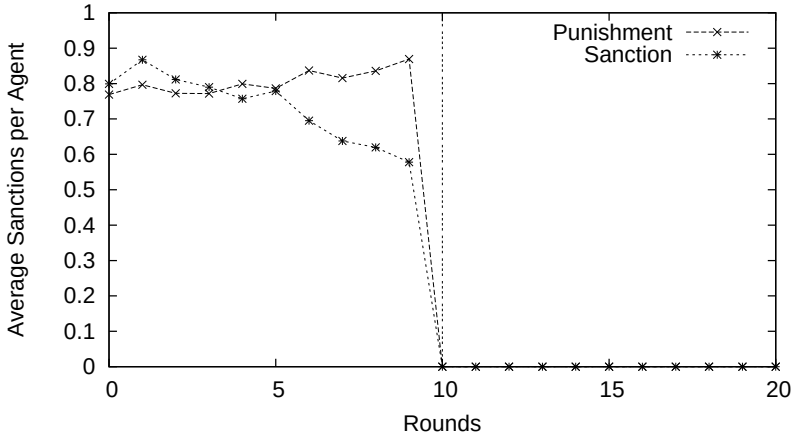
The dynamics of punishment and sanction (shown in Figure 7.3(b)) are also similar after this adaptation time, being the average amount of punishments and sanctions similar at the end of the punishment block wrt the humans experiment. We observed that, in both humans and simulated agents, the number of sanctions is lower than the number of punishments, representing a cheaper cost for society to obtain cooperation.

These experiments have been useful to test the correct implementation of the EMIL-I-A architecture and a satisfactory selection of the Salience Weights.

[5] We want to remind the reader that we consider a sanction to be a punishment with a normative message.



(a) Average Cooperation



(b) Average Punishment/Sanction

Figure 7.3: Punishment and Sanction Experiment with EMIL-I-A Simulated Subjects

Player 1 \ Player 2	Cooperate	Defect
Cooperate	3 3	5 0
Defect	0 5	1 1

Figure 7.4: Prisoner's Dilemma Game Payoff Matrix

7.3 Exploiting EMIL-I-A

So far we have seen how EMIL-I-A and the Salience Control Mechanism have been correctly designed, obtaining similar results when comparing its behaviour with that of human subjects. To perform an exhaustive analysis of these two mechanisms, we introduce a simulation model where to test them.

7.3.1 Simulation Model

In order to capture the specific dynamics of punishment and sanction and to test their relative effects in the achievement and maintenance of cooperation a simulation model has been developed.

In this model, agents play a variation of the classic Prisoner's Dilemma Game (PDG) (whose payoff matrix is shown in Table 7.4), where an extra stage has been included in the game: after deciding whether to cooperate or not, agents can also choose whether they want (or not) to punish or sanction the opponents who defected. The motivation behind the PDG is due to our long-term research goal aimed at studying enforcing technologies in virtual societies, and more specifically in environments like P2P scenarios or web-services markets. These types of scenarios share a number of characteristics with the PDG: dyadic encounters, repeated interactions, and one-shot games.

Each timestep of the simulation is structured in 4 phases, that are repeated for a fixed number of timesteps:

1. **Partner Selection**: Agents are randomly paired with their neighbours in the social network.

2. **First Stage**: Agents play the PDG, deciding whether to cooperate (C) or to defect (D).

3. **Second Stage**: Agents decide whether to punish/sanction or not the opponents who defected. Only agents who have recognized that there is a norm of cooperation governing their group use sanction to enforce others' behaviours; otherwise punishment is used. Punishment works by imposing a cost to the defector, this way affecting its payoffs. On the other hand, sanction also informs the target (and

possibly the audience) that the performed action violated a social norm, thus having an impact both on agents' payoffs and on the process of norm recognition and norm salience. If an agent decides not to punish/sanction and it is a norm-holder (i.e. an agent with a highly salient norm of cooperation stored in its mind), it can send an educational message to its opponent. Problems related to the way agents communicate the abstract concept of norm are outside of the scope of this paper and therefore, we assume perfect communication.

4. **Strategy Update**: As agents have mixed strategies, these strategies are updated on the basis of their decisions, payoffs and social information acquired.

7.3.2 Experimental Design

One of the main objectives of this research is to study the achievement of cooperation in adverse situations, where defecting is the utility-maximizing strategy for the agents. In order to observe the relative effects of punishment and sanction in our artificial scenario, we exploit the advantages of having agents endowed with normative minds (fine-tuned with the parameters obtained with human-subjects in similar scenarios) allowing them to process the signals produced by these two enforcing mechanisms separately.

In order to analyze the differences of both types of punishment on the agents' decision making, we have performed an exhaustive experimental analysis. To reduce the search space (and save computation costs), we have prefixed the population size to 100 agents.

In the following experimental section, we contrast the results obtained in simulations where only punishment is allowed against situations in which both punishment and sanction are allowed. Only after having recognized the existence of the cooperation norm, agents will sanction defectors, thus defending the norm; otherwise they will just punish them.

As in this work we are not interested in analysing the emergence of norms, some agents already endowed with the cooperation norm are initially loaded into the simulation: we refer to them as *initial norm's holders (INHs)*, and are initially loaded with the norm at a salience of $0,8$. If no agent had the norm, they would have to start a process of norm emergence that would include the recognition of an anti-social behavior, the identification of a possible solution and the consequent implementation in the society in the form of norms.

All the following experiments have been run on two different topologies: a fully connected network (where agents have access to complete information) and a scale-free network (where agents only access local information). Despite obtaining the same final convergence results in both topologies, we have observed a delay in the scale-free networks, caused by the cascade effect of the spread of norms produced by the location of the INHs (better connected INHs produce faster convergence results).

7.3.3 Punishment and Sanctioning on the Emergence of Cooperation

In this first experiment, we analyze the relative effects of punishment and sanction on the achievement of cooperation, paying attention to the amount of INHs and the different damages imposed with punishment and sanction. By comparing the results obtained with different damages in Fig. 7.5, we observe that different damages (i.e. the amount of punishment/sanction imposed to the target) affect the cooperation levels differently. We can understand this first experiment as a proof-of-concept to check the correctness of our architecture. As expected, agents' motivation to defect decreases in a much stronger way with a damage of 5 than with a lower damage of 3[6]. It has to be pointed out that, despite what happens when using punishment, in populations enforced by sanction (with a minimum amount of INHs, 10 in this experiment), cooperation is also achieved when imposing a lower damage, as can be seen in Fig. 7.5(a), Fig. 7.5(c), and Fig. 7.5(e).

Both punishment and sanction directly affect the agents' SID, reducing their motivation to defect. However, these two enforcing mechanisms are effective in achieving deterrence only when the damage imposed is at least 3, as with a damage of 3 (Cooperation Payoff = 3, Defection Payoff = 5 - 3 = 2) cooperation is the utility-maximizing strategy.

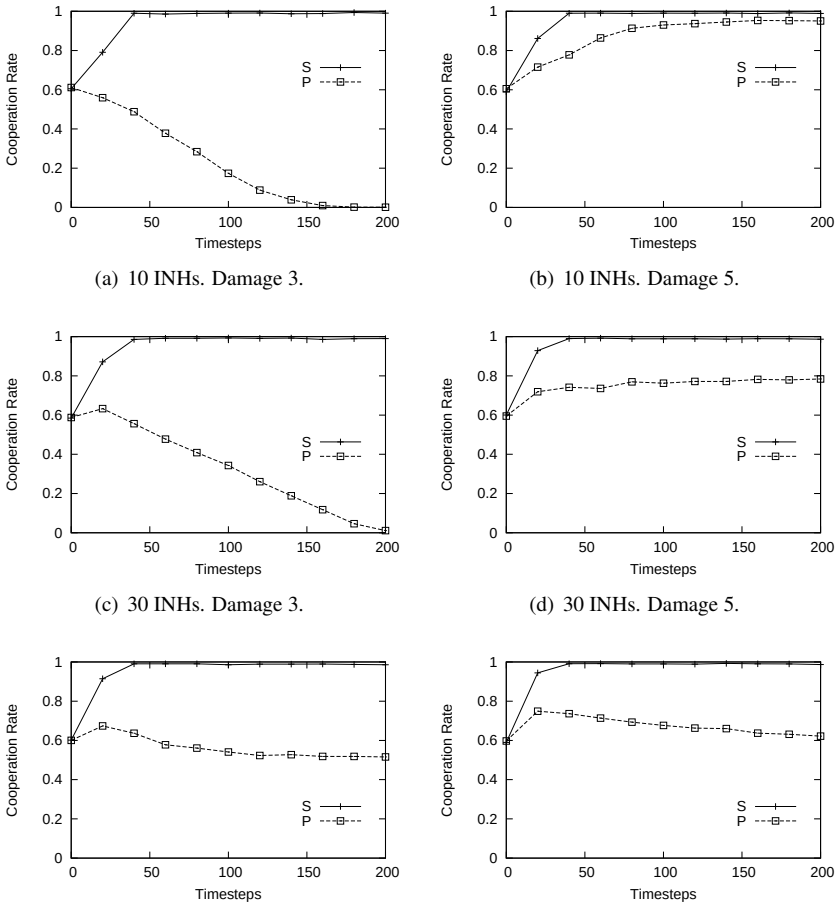
As said in Sec. 7.1.1, agents' cooperation probability is affected by both the SID and the ND: sanction - thanks to its signalling component - influences the normative drive more than punishment. In order to obtain deterrence, punishment exploits the power of norms much less than sanction, that is why it needs to impose higher damages on its targets.

The amount of INHs produces an interesting result when using sanctioning: when the number of INHs is increased, emergence is achieved faster and it follows a distribution equal to the neighbors distribution (in the regular networks the emergence of cooperation increases linearly and exponentially in scale-free networks). When using punishment the dynamics of the cooperation are different, as the salience is not affected as strongly, and therefore the normative drive does not push towards cooperation as much as when sanction is used. In the next section we analyze this phenomenon.

7.3.4 Norms Spreading

With the proof-of-concept experiment presented in the last section, we not only tested the effects of punishment and sanctioning on the achievement of cooperation, but also, how the initial amount of norm holders affects the dynamics of cooperation. Analysing the same experimental data presented in the previous section, we now observe the relative effects of sanctions and punishments on the recognition of the cooperation norm. As specified previously, our hypothesis is that, thanks to its signaling nature, sanction

[6]These values have been chosen for experimentation as both 3 and 5 punishment damages turn the cooperative action into the utility-maximizing strategy. A damage of 3 produces a *slight* improvement for cooperation (Payoff = 3) over the defection (Payoff 5 - 3 = 2). On the other hand, a damage of 5 produces a *stronger* difference between cooperation (Payoff = 3) and defection (Payoff = 0).



(a) 10 INHs. Damage 3.

(b) 10 INHs. Damage 5.

(c) 30 INHs. Damage 3.

(d) 30 INHs. Damage 5.

(e) 50 INHs. Damage 3.

(f) 50 INHs. Damage 5.

Figure 7.5: Effects of Punishment (P) and Sanction (S) on the Emergence of Cooperation.

allows norms to spread more and more quickly in the population, making it more resilient to change than if enforced only by mere punishment.

By paying attention to the average norm's salience per agent in Fig. 7.6, we appreciate that sanction produces a stronger effect on the spread of norms than punishment, thus verifying our hypothesis. This phenomenon is mainly caused by the design of our normative architecture, where sanctions impact in a stronger way the agents' recognition and salience of norms.

Moreover, this is a self-reinforcing process: once an agent has recognized a norm, it will start sanctioning, getting others to comply with the norm, reproaching transgressors and reminding would-be violators that they are doing something wrong. This process affects and reinforces the sanctioner's own salience, but also the norm's salience of its neighbours.

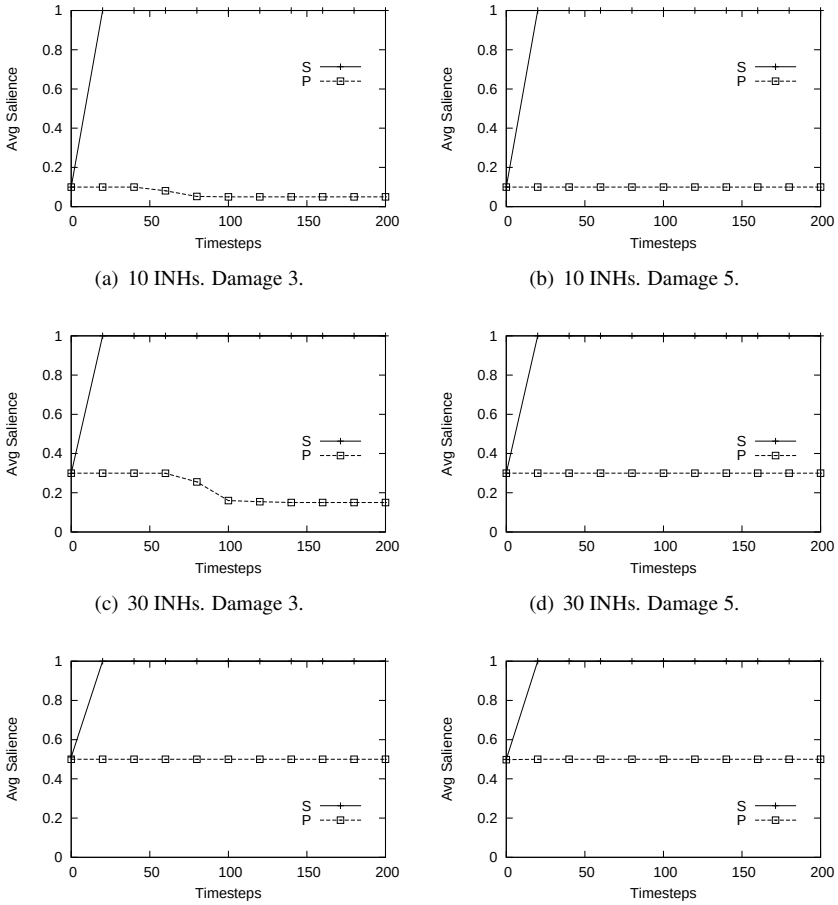
7.3.5 Costs of Punishment

Our second hypothesis is that sanctioning combines high efficacy in discouraging defectors with lower costs for society as compared to punishment. From the data shown in Table 7.1, we observe that in the two situations where both punishment and sanction allow cooperation to emerge (i.e. imposing a damage of 5), sanctions are 32,52% better in terms of the amount of punishment/sanction occurrences and the cost for the society wrt the situation where punishments are used. In other words, when using sanction, the amount of punishment/sanctioning acts and consequently the associated costs are decreased by 1/3. This is an interesting result that confirms our hypothesis, that the use of sanctions reduces the social costs of the achivement of cooperation as it impacts deeper on the salience.

	Punishments	**Sanctions**	**Individual Cost**	**Social Cost**
Punishment	4,8748	0	5	24,374
Sanction	0,3364	1,2491	5	7,9275

Table 7.1: Costs of Punishment and Sanctioning. Average values per agent and timestep.

Moreover, (and even though the data obtained with Damage 3 are not comparable amongst them because of the different effects on the emergence of cooperation), we can observe that - from a system designer point of view - knowing the right value of the punishment/sanction to apply (5 better than 3 in this case) would reduce the global costs, obtaining the same result on the emergence of cooperation (full cooperation). The reason for this phenomena is because a sanction of 5 would produce a stronger deterrent effect in the self-utilitarian drive of the agents than a sanction of 3, which needs to be repeatedly applied to achieve the same results.



(a) 10 INHs. Damage 3.

(b) 10 INHs. Damage 5.

(c) 30 INHs. Damage 3.

(d) 30 INHs. Damage 5.

(e) 50 INHs. Damage 3.

(f) 50 INHs. Damage 5.

Figure 7.6: Effects of Punishment (P) and Sanction (S) on the Salience.

7.3.6 Dynamic Adaptation Heuristic

In the experiment in Sec. 7.3.3, the damage that both punishers and sanctioners can impose on defectors in order to deter them from violating again is fixed for the entire duration of the simulation. But in some circumstances, this damage could be reduced, without lowering its deterring effect. In order to allow our agents to dynamically choose the right amount of punishment and sanction to impose, thus reducing social costs, a *Dynamic Adaptation Heuristic (DAH)* has been implemented. This heuristic works in the following way:

1 **if** $Defectors_{t-1} \leq Defectors_t$ *AND* $Defectors_t > ToleranceDef$ **then**
2 Increase PunCost by Δ;
3 **if** $Defectors_{t-1} > Defectors_t$ *OR* $Defectors_t < ToleranceDef$ **then**
4 Decrease PunCost by Δ;

Algorithm 4: Dynamic Adaptation Heuristic (DAH).

By keeping track of the amount of defectors in the previous timestep ($Defectors_{t-1}$), and comparing it with the actual amount of defectors, the imposed cost of punishment (PunCost) is adapted consequently with a Δ ($\Delta = 0.1$ in this work), thus obtaining an intelligent dynamic adaptation. $ToleranceDef$ is another parameter used to allow agents to be tolerant against exploration from the other agents, allowing a percentage of violations in their social environment. If $ToleranceDef$ is fixed to zero, it might produce a continuous rise of the Punishment Cost without having an effect on the violators mind (as some violations are done by mere exploration).

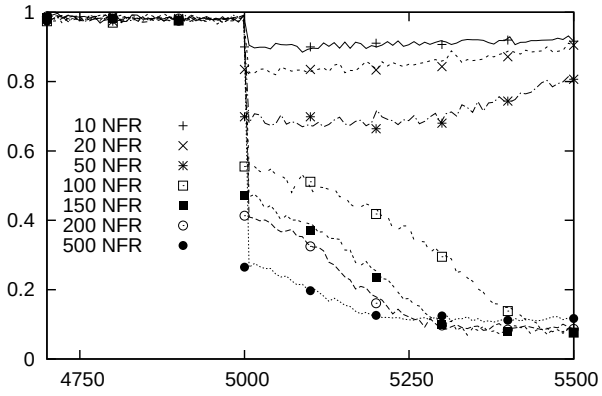
	Punishments	**Sanctions**	**Individual Cost**	**Social Cost**
DAH Punishment	2,1142	0	5,1640	10,9177
DAH Sanction	**0,2932**	**0,8669**	**3,7031**	**4,2959**

Table 7.2: Performance of the Dynamic Adaptation Heuristic. Average values per timestep.

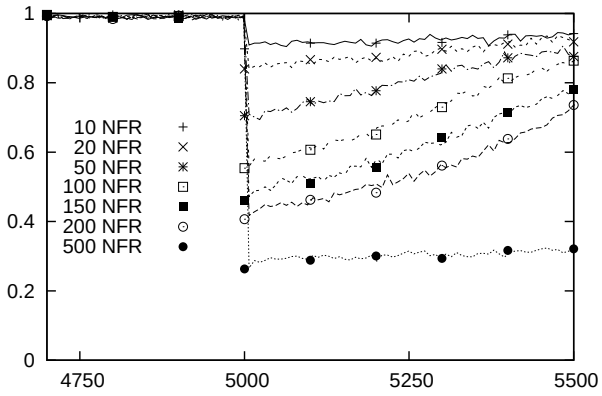
In order to test how the implemented DAH affects the performances of both punishment and sanction, an exhaustive exploration of the search space with different amounts of INHs (from 10 to 90, increasing at intervals of 10) has been performed.

Results obtained for different amounts of INHs follow the same distribution, for the cases with and without DAH. In Table 7.2, the average performances of both punishment and sanction in the dynamic and static conditions are shown. The DAH allows society to significantly reduce the number of punishing (57% less) and sanctioning (27% less) acts with respect to the static condition reported in Table 7.1.

However, when using DAH, the average individual cost of punishment is slightly higher than that of sanctioning. This is given by the cyclic dynamics produced by agents driven only by their strategic drives. When enforced by punishment, agents abandon their defecting strategy because they want to avoid the cost of punishment. The number of cooperators starts to increase and punishers decrease the punishment cost accordingly. However, this reduced punishment cost makes defection to become



(a) Punishment.



(b) Sanction.

Figure 7.7: New Free Riders introduced at TS = 5000.

the utility maximizing strategy. Consequently, the number of defectors will increase again and the punishment cost accordingly. With this newly adapted punishment cost, defection will not be the optimal strategy anymore, reaching the initial situation again.

Consequently, the social expenses using DAH are considerably reduced both when punishing (66%) and sanctioning (56%). The implemented heuristic allows society to intelligently reduce the social costs needed for the achievement and maintenance of cooperation.

7.3.7 Adapting to the Environment: Free Riders Invasions

This experiment is aimed to test the hypothesis that sanction makes the population more resilient to change than if it were enforced only by mere punishment. If suddenly a

large amount of new defectors joins the population, we suppose that defectors will take longer to invade the population in which sanction has been used. In order to confront the relative speed of adaptation and degree of resilience of the populations enforced with punishment and sanction, we run simulation experiments in which after timestep 5000, new free riders (from a minimum amount of 10 to a maximum of 500) are injected in the populations

Experimental results (see Fig. 7.7) show that a population enforced by (DAH) sanction is able to receive up to 200 new free riders and still to maintain a high level of cooperation, while when (DAH) punishment is used, only 100 new free riders make cooperation collapse. In the population enforced by sanction, a larger amount of cooperation norms have spread, this having a refraining effect on the decision of abandoning the cooperative strategy. Highly salient norms guarantee a sort of inertia, restraining agents to change their strategy to a more favorable one.

7.4 Conclusions

The Distributed Punishment experiment performed with human subjects (presented in Chapter 5) gave us the hint of the existence of a heavier "message" in certain types of punishment, affecting differently the reasoning of subjects. The designed EMIL-I-A Architecture takes into consideration the difference stated by cognitive scientist [Giardini et al., 2010] between *Punishment* and *Sanction*. We rebuilt the Distributed Punishment experiment in a simulated scenario with EMIL-I-A agents, obtaining satisfactory results recreating with an virtual agent the behavior of human subjects under the same experimental conditions.

Meanwhile, we have also proven with experimental economics the hypothesized difference between punishment and sanction. As far as we know, this is the first time this result has been proven, being therefore interesting for such community. On the other hand, EMIL-I-A has also performed well recreating the human subjects dynamics in the same experiment. After proving the validity of such architecture, we (computer scientists and policy makers) are given with a powerful tool that allows us to further exploit these punishment technologies, in different situations, unfeasible to obtain in the laboratory with human subjects.

For that reason, we have developed a simulation platform that have served us to test our normative theory of punishment and the designed EMIL-I-A architecture. The simulation results obtained by exploiting the capabilities of EMIL-I-A show the ways in which punishment and sanction affects the emergence of cooperation. More specifically, these results seem to verify our hypotheses that the signaling component of sanction allows this mechanism (a) to be more effective in the achievement of cooperation; (b) to spread norm faster and wider in the population, making it more resilient to environmental change than if enforced only by mere punishment; (c) to reduce significantly the social cost for cooperation to emerge.

Moreover, and up to our knowledge, this is the first work in which agents are endowed with rich cognitive architecture allowing them to be affected by the normative information associated to sanction and to dynamically gauge the amount of damage to impose on defectors, achieving an efficient compliance.

However, one question remains unanswered: what happens when punishment is not available but the norm is still necessary? We need a special mechanism that allow agents to be compliant with the norms independently of the external outcomes. Axelrod defined this process as internalization [Axelrod, 1986] and we will show in the following chapter how it is tightly linked with the salience of norms. By donating agents with norm internalization mechanisms we will achieve a more compliant, stable and cost-efficient society, without sacrificing adaptability.

Chapter 8

Internalization

In the previous chapter we have seen how punishment is a useful mechanism to impose social norms by deterring (and educating) norm violators. However, we have seen how in the absence of punishment mechanisms, agents tend to adapt to the self-interested utility-maximizing strategy.

The problem social scientists still revolve around is how autonomous systems, like living beings, perform positive behaviors toward one another and comply with existing norms, especially since self-regarding agents are much better-off than other-regarding agents at within-group competition. Since Durkheim, the key to solving the puzzle is found in the theory of internalization of norms [Mead, 1963, Parsons, 1937, Grusec and Kuczynski, 1997, Gintis, 2004, Horne, 2003]. One plausible explanation of voluntary non self-interested compliance with social norms is that norms have been internalized. Internalization is another of the mechanisms identified by Axelrod [Axelrod, 1986] for the establishment of norms[1].

Internalization occurs when

> *a norm's maintenance has become independent of external outcomes - that is, to the extent that its reinforcing consequences are internally mediated, without the support of external events such as rewards or punishment* ([Aronfreed, 1968, p 18]).

This process entails several advantages. For example, norm compliance is expected to be *more robust* when norms are internalized than in the case when norms are external reasons for conduct: if everybody in the population internalizes a norm, there is no incentive to defect and the norm remains stable [Gintis, 2004]. Driven by internal motivations, internalizers are not only much better at complying with norms than are externally-enforced individuals, but also at defending them. An effect of the latter prediction is that norm internalization is decisive, if not indispensable, for distributed social control. Internalization is not only a mechanism of private compliance, but also a factor of *social enforcement*. Individuals who have internalized the norm, not only

[1] Other factors identified by Axelrod in [Axelrod, 1986] like dominance and deterrence have been analyzed in Chapter 7.

comply with it without any need of external enforcement, but in many circumstances they also want others to observe the norm, by reproaching transgressors and reminding would-be violators that they are doing something wrong.

The importance of norm internalization in favouring social order has been largely recognized, but philosophers, social scientists, psychologists, anthropologists still strive to answer some fundamental questions: Why and how do agents internalize social inputs, such as commands, values, norms and tastes transforming them into endogenous states? Which types of mental properties and ingredients ought individuals to possess in order to exhibit different forms of compliance? How many people must internalize a norm in order for it to spread and remain stable? What are the different implications of distinct modalities of norm compliance for society and governance of distinct modalities of norm compliance? These questions received no conclusive answer so far. In particular, no explicit, controllable, and reproducible model of the process of internalization is available yet.

Agents conform to an internal norm because so doing is an *end* in itself, and not merely because of external sanctions, such as material rewards or punishment.

This chapter provides an operational model of the internal mechanisms of internalization. Rather than a none-or-all phenomenon, norm internalization is a highly dynamic process, whose deepest step occurs when norms are observed thoughtlessly. Norm internalization is here characterized as a *multi-step* process, occurring at various levels of depth and giving rise to more or less robust compliance and also allowing in certain circumstances for automatic behaviours to be unblocked and deliberate normative reasoning to be restored. If only the deepest level of internalization were modelled, it would be sufficient to hardwire agents with norms and make them automatically compliant. Even though in several contexts this type of agent is highly efficient, in rapidly changing environments it proves to be inadequate to deal with the adaptive problems it faces. Thus, to fully operationalise such a multilevel model of norm internalization requires a complex agent architecture. Unlike the vast majority of simulation models in which heterogeneous agents interact according to simple local rules, in our model the agents (EMIL-I-As) are provided with normative mental modules, allowing them to internalize norms active in the social system in which they interact.

Emotions, playing a significant role in this process, will not be investigated at this stage. However, recent work [de Melo et al., 2011] has explored the effects of emotions in hybrid environments (populated with humans and agents) affecting on the decision making of human subjects.

8.1 What is Internalization?

Norm internalization is the process by means of which norm's compliance becomes independent of external outcomes i.e. when the norm addressee observes it free from external punishment and reward. In other words, it is a mental process that takes a (social) norm as input and provides the individual with terminal goals, i.e. goals that are considered as an end in itself and as means for achieving other goals. Therefore, it is the next step to take in our road towards norm compliance after having studied the punishment/sanction mechanisms.

As discussed in Sec. 6.2, normative beliefs generate normative goals by reference to an external enforcement. However, a norm is internalized when the norm addressee complies with it independently of external sanctions and rewards. In such a case, the normative goal is *no more relativized* to an expected sanction, but only to the main normative belief.

8.1.1 Factors affecting internalization.

Why do agents observe a norm irrespective of external enforcement? Far from providing a complete list of the factors favouring norm internalization, in the present work we will focus on some of the elements playing a key role in this process, such as:

- consistency;
- self-enhancing effect;
- urgency;
- calculation cost saving;
- norm salience.

Let us start with *consistency*. This mechanism operates at two stages: first by selecting which norm to internalize, and later by enforcing it (self-enforcement) and controlling that one's behavior corresponds to it (self-control). Consistency of new norms with one 's beliefs, goals and previously internalized norms plays a crucial role in the selection process (of which norm to internalize). Successful educational strategies favor internalization processes, often by linking new inputs with previously internalized norms. Analogous considerations apply to policy-making. Consider the antismoking legislation: the efficacy of antismoking campaigns based on frightening announcements and warning labels (e.g., sentences like 'Smoking kills' on cigarette packages, see [Goodall, 2005]) is still controversial. One of the factors reducing the efficacy is the effect known as *hyperbolic discounting* ([Bickel and Johnson, 2003]; see also, [Rachlin, 2000]), a psychological mechanism that leads to invest in goal-pursuit a measure of effort that is a hyperbolically decreasing function of the time-distance from goal-attainment, and leads people to procrastinate energy-consuming work until the very last moment. Due to hyperbolic discounting, people, especially young people, are unable to act under the representation of delayed consequences of current actions. On the other hand, much more efficacious anti-smoking campaigns are those playing on previously emerged and diffused set of social norms, such as the live-healthy precepts, highly consistent with the message they want to transmit.

The second factor playing a role in norm internalization is the *self-enhancing* effect of norm compliance: the norm addressee realizes that it achieves one of its goals by observing a given norm. Suppose I succeed in refraining from smoking and that after a few days, I realize an advantage that I had not perceived before: food starts to taste again. This discovery generates a goal (quit smoking to enjoy good food), not relativized to the norm but supporting it: I have converted the norm into an ordinary goal. Whether this goal will be strong enough to out-compete addiction is another matter.

Third, we focus on *urgency* (see, for example, [Zhang and Huang, 2006]). In particular, one can argue that the more a given norm allows to answer problems frequently encountered under conditions of urgency, when time for decision-making is none or scanty, the more likely that norm will be internalized.

Fourth, we claim that agents are *parsimonious calculators*: under certain conditions, they internalize norms in order to save calculation and execution time. Upholding a norm that has led one to succeed reasonably well in the past is a way of economizing on the calculation costs that one would have to sustain whenever facing a new situation. Imagine a driver's decision to stop at the traffic light when it turns red. Each time our driver approaches a red traffic light, it calculates the costs and benefits of complying with the norm: e.g. it predicts that it will save time by ignoring the norm, but that a fine will then follow with a certain probability. The driver then chooses what is best for itself.

After a certain amount of calcula, always returning the same output (e.g. the driver always deciding to stop in order to avoid punishment), the agent able to internalize will abstain from instrumental reasoning: it will stop the car when traffic light is red without calculating the convenience of such behavior. After having weighted the costs and benefits of complying or not with a certain norm for a certain number of times (and having reached exactly the same decision every time), the agent stops calculating and takes norm compliance as the best choice. By doing so, it will save time acting in a more efficient way. Explicit processing starts declining with norms gaining force, and gradually stops once norms have been internalized.

This last point is strictly intertwined with another important factor favoring norm internalization: i.e. norm salience.

Norm's salience (already analyzed in Sec. 6.3) is another factor allowing norm internalization: if the norm is highly salient, the norm is a candidate to be internalized by the agent. We claim that social norms are not static objects and the degree to which they are operative and active varies from group to group and from one agent to another: we refer to the degree of activation as *salience*. The more salient a norm is, the more it will elicit a normative behaviour [Bicchieri, 2006, Xiao and Houser, 2005, Cialdini et al., 1990]. Then, salience may increase to the point that the norm becomes internalized, i.e. converted into an ordinary goal, or even in an automated conditioned action, a routine. Norm's salience is a complex function, depending on several social and individual factors.

The salience of a normative belief can vary depending on several social and individual factors. The social factors affecting salience have been discussed in Chapter 6. On the other hand, norm salience is also affected by the individual sphere, it depends on the degree of entrenchment with beliefs, goals, values and previously internalized norms[2].

In this work we will focus only on the last two conditions, *calculation cost saving* and *norm salience*, leaving the rest open for future work.

[2]It has to be pointed out that the norm salience can also gradually decrease: for example, this happens when agents realize that norm violations received no punishment or when normative beliefs stay inactive for a certain time lag, suggesting that the norm is not very active in the population anymore.

In the next sections, we will show how EMIL-I-As can account for *reversible routines*, or, which is the same, for flexible conformity.

8.2 Dynamics of Norms Internalization

When located in environments regulated by different competing norms, agents need a mechanism to detect the respective importance of each norm, thus allowing them to select which one of them to internalize, when to internalize it, or to de-internalize one norm in order to internalize another. EMIL-I-A is given with several modules to handle this complex task. Some of these modules (like the *norm recognition module* and the *salience control module*) have already been explained in Sec. 6.3. Their structure and functioning remain completely the same, and in this section we will center in the *Internalization Module*. This module is a conceptual selector of the decision making function to be followed; agents can follow an instantaneous utility maximization function, a normative decision function or an automatic response (with no computation associated).

8.2.1 Internalization Module

The internalization mechanism accounts for the trade-off between self-adaptation and internalization. This is in charge of selecting which norm should be internalized, converting it into an automatic behavior that is fired without any decision making process. Of course, automatization is a multi-step complex process that needs conceptual, modelling and experimental work. For the time being, we implement norm compliance as either automated or decided upon behaviour. In the real matters, the alternative is probably fuzzier, and different steps on the continuum from a fully deliberate to a wholly automated behaviour ought to be modelled. In the present work, (and as we can see in Figure 8.1) an internalized norm fires an automated action without any decision making process. An agent can internalize only one norm regulating a certain situation (e.g. people can greet by waving from a respectful distance, with a friendly handshake, or with a number of kisses: depending on the environment, only one of these actions will be internalized). Moreover, thanks to the information provided by the Salience Control Module, this mechanism is also in charge for the de-internalization process. As explained in Sec. 8.1.1, EMIL-I-As are *parsimonious calculators*, and therefore, they internalize norms in order to save calculation (decision-making) and execution time. Trying to record the monotony of agents' decisions, agents keep track ($Eval^N$) of the amount of times this decision-making calculation (in which the expected payoffs obtained by complying or not with the norm is calculated) is performed and it returns the same decision. In case the decision-making calculation would return a different decision, the counter would be resetted, representing a break up of the monotony.

- Two conditions are necessary for a norm N to be internalized by an agent:

    1. the salience of the candidate norm is at its maximum value ($Sal_t^N = 1$), and,

2. the decision-making calculation has been done for a certain amount of time always returning the same decision (in the present model we fixed the calculation repetition tolerance to 10, therefore $Eval^N \geq 10$).

- Instead, a norm $N1$ is de-internalized when the following conditions apply:

1. the salience of the internalized norm arrives at its minimum value ($Sal_t^{N1} = 0$), and,

2. the salience of another norm (ruling exactly the same situation) exceeds the internalized one ($Sal_t^{N2} > Sal_t^{N1}$)

Summing up, we can see in Figure 8.1 a representation of the EMIL-I-A architecture with all its capabilities. An example decision making process would start recognizing the situation in which an agent finds itself; after that an agent reacts with the automated action in case a norm is internalized. Otherwise, and if the agent has not recognized any norm, it will use the *Norm Recognition Module* and then proceed with a default decision-making. However, if the norm was already recognized, but not yet internalized, agents check the internalization conditions previously explained and check whether the norm can be internalized; the decision making at this stage will be that restricted to the normative beliefs and goals, orchestrated by the salience.

8.2.2 Urgency Management Module

In self-organizing societies, where agents are in charge of creating, imposing, modifying, and enforcing norms, these are neither explicitly specified nor defended by any in-charge authority. In a number of situations norms must be broken and agents should be able to recognize these situations properly. By this means, agents have to intelligently decide to violate the norm, for example crossing at the red light when hearing an ambulance coming from behind.

In our model, to violate norms when urgent situations occur, agents need to explicitly record and compare the urgent and necessary nature of the situation their are facing (in the ambulance example, the lights and the siren make explicit the socially desirable violation of *Do not cross in red* norm). Therefore, all the *Urgent Situations* are accompanied with a salience value ($Sal_t^{Urgency}$), that agents can compare with the salience of the norm "normally" regulating that specific situation and decide accordingly. We are aware of the fact that associating *Urgent Situations* with a pre-defined salience value is a short-cut, and agents should learn this Urgency salience instead, but at the moment how agents learn to recognize what it is considered to be an urgency has been deliberately left out and will be explored in future work. Therefore, in urgent situations the decision making compares the salience of the urgent situation with the salience of the norm that usually regulates the situation, as it is showed in the following algorithm:

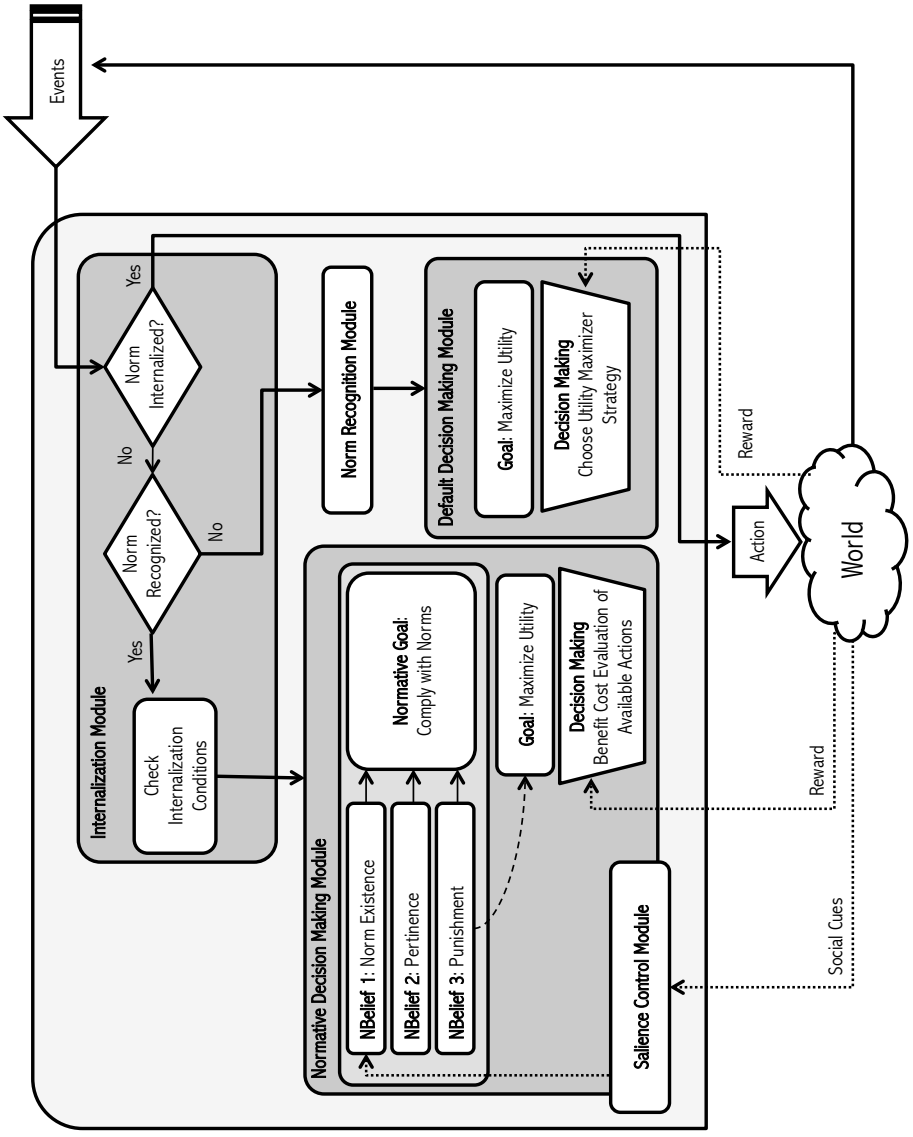


Figure 8.1: EMIL-I-A Architectural Design

```
1 if Sal_t^N ≤ Sal_t^Urgency then
2 |   Violate the norm;
3 else
4 |   Comply with the norm;
5 end
```

Algorithm 5: Urgency Management Algorithm.

8.3 Self-Regulated Distributed Web Service Provisioning

Our simulation scenario consists in a web-service market populated by agents whose task is to find out which services are offered by other agents (distributed) and control the behavior of the rest of peers (self-regulated). This environment allows for EMIL-I-A to be tested in a setting relevant for the MAS community. We will use the complete EMIL-I-A architecture, although with a simplified version of the Decision Making than the one shown in Sec. 7.1.1. The decision making of the agents used in here are explained in Sec. 8.3.4.

The simulated scenario presents some fundamental features reproducing the real-world phenomenon it models: it is dynamic (new services with different capabilities can be created during the simulation), unpredictable and populated by heterogeneous agents.

To test the performance of internalizers in a multi-norm scenario, we have introduced two possible norms regulating the system. This variant is a novel contribution with respect to previous works in the normative self-adapting community (such as [de Pinninck et al., 2007b, Blanc et al., 2005]).

8.3.1 Motivation

The presented simulation model aims to reproduce a self-regulated web-service market. Inspired to the "Tragedy of the Digital Commons", we present a provider-consumer scenario, populated by providers that might suffer a tragedy-of-the-commons caused by the consumers' over-exploitation. The services we refer to present the following features: (1) they are available for use, and return agents a certain utility; (2) they are a common good, and, (3) they can attend up to a certain amount of requests: once a certain threshold is overcome the service starts loosing quality.

While looking for services, consumers ask central registries information about who can provide them the service they are looking for, information that is obtained by mobilizing neighbors in their social network. Networks are an effective way to obtain information - for example provided through word-of-mouth - and represent an alternative source, with respect to traditional methods. Conventional approaches in multi-agent systems, such as *registries* (centralized repositories with the services that are offered) or *matchmakers* (auxiliary agents with knowledge to couple agents offering services and agents with need of services), partially address this problem [Decker et al., 1997].

However, in highly dynamic environments, there is a valuable amount of information that cannot be stored in a centralized repository. In some cases, much of this information (such as the updated details about the quality of the service or the availability of the service) may be accessed only by using social networks of interaction. In this work, a hybrid approach has been proposed: similarly to the *white pages* in UDDI [Curbera et al., 2002], our agents will query a central server for obtaining pointers to service providers, and then, all the other important information about the service will be provided by the service providers. The functioning of the system is similar to that implemented by Napster [nap, 2006].

More specifically, our case study presents the following dynamics: during the simulation, agents look for and find services (or learn tasks) that are necessary to satisfy their necessities. According to the number of agents that are exploiting it, every service has a level of quality that can increase or decrease during the simulation. Agents obtain the services they need by finding other agents willing to share the service they have control of. For the sake of simplicity, the allocation of new resources and necessities are done automatically from the system into the agents, with different allocation probability distributions depending on the experiment.

By finding a needed service with its expected quality, an agent obtains a reward; however if the service obtained does not fulfil the agent's expectations - in terms of quality -, the requester will continue to look for the service. The service itself does not take any part in the decision to be shared or not: the decision is taken by the service holder, and, if a service-holder decides to share its service, this will be automatically shared.

Moreover, when sharing, a service-provider remains blocked for a number of time-steps (representing the *transaction* time), thus reducing its possibility to look for and find its own needed service. This *transaction* time is inspired on the P2P networks: in these type of networks, when a resource is shared, the bandwidth of both agents is reduced (one for uploading the resource and the other for downloading it).

Thus, we can see how in this scenario, a selfish agent is motivated to free-ride by obtaining services from other agents, but giving none. The spread of this selfish behaviour would eventually lead to a "tragedy of the commons" situation, getting the society to collapse, as predicted in [Adar and Huberman, 2000].

Trying to model the performance of a real system, EMIL-I-As behave dynamically, changing the rates of services' assignation and request along the time. To test how the *EMIL-I-A* architecture performs when facing this "tragedy of the digital commons" situation, we compared its performance with that of agents endowed with different architectures.

8.3.2 Norms in our Web Service Scenario

As we have already specified, the service providers can share a resource with a requester. However, the service provider is the only one that knows at run-time the quality of the service he can offer, and the quality needed by the requester. We introduced two different norms that regulate the environment: (N1) the *Always Share when Asked for a Service* norm and (N2) the *Share Only High-Quality Services* norm. Depending on the environmental conditions (the service's capacity distributions) and on the agents'

necessities (the services' expected quality), one or the other norm will become more salient, and therefore will govern the behavior of the system.

In order to enforce norms and to deter agents from violating them, EMIL-I-As can use the two different enforcing mechanisms already analyzed: punishment or sanction.

8.3.3 Model

For the sake of clarity, we model the environment as a *Hybrid Decentralized* architecture (as in Fig. 8.2), as defined in [Androutsellis-Theotokis and Spinellis, 2004]. In this type of architecture, each agent shares services (or resources) with the rest of the network. All agents are connected with a central directory server that maintains (a) a table in which all the users' information are recorded (normally the IP addresses and the connection bandwidth, which in our system are represented as an ID and the capacity of the service), and (b) a table listing the services from each agent, along with metadata descriptions of the services (such as the type, capacity and so on). Every agent looking for a service sends a query to the central server. The server searches for matches in its index, returning a list of users that hold the matching service. The user can then open direct connections with one or more of the peers that hold the requested service, and he can request the needed service. The final decision of whether sharing or not the service is to be taken by the peers.

In our scenario, the central server maintains the following information: A, the set of agents in the system, SN, the social network that connects the agents, and T, the types of services that can be provided and looked for in the system.

Agents are located in a social network that restricts their interactions and communications. The social network is described in the following way:

$SN = \{A, Rel\}$

where A is the set of agents populating the network and Rel is a neighborhood function. The neighborhood function is indirected, therefore is $\forall a,b \in A$ if $Rel(a,b)$ then $Rel(b,a)$.

During the simulation, agents find resources (or learn tasks) that are offered as services:

$Service = \{ID, t, c, Clients\}$

where ID represents the identification number of the service provider (e.g. IP address), $t \in T$ is the type of service offered by the service provider (e.g. storage, calculation time, data analysis), c is the capacity of the service provider, i.e. the amount of clients it can satisfy offering a good Quality of Service; and $Clients \subset A$ indicates all the clients the service provider is holding at the moment.

Moreover, agents find out new necessities, that can be fulfilled by obtaining services from other agent:

$Needs = \{t, q, d\}$

where $t \in T$ represents the type of service an agent needs with a minimum q Quality of Service level, before timestep d. The Quality-of-Service of a provider is calculated in the following way: $\frac{Capacity}{Clients}$

For the sake of simplicity, the allocation of new resources and necessities to the agents is done automatically by the system, with different allocation probability distri-

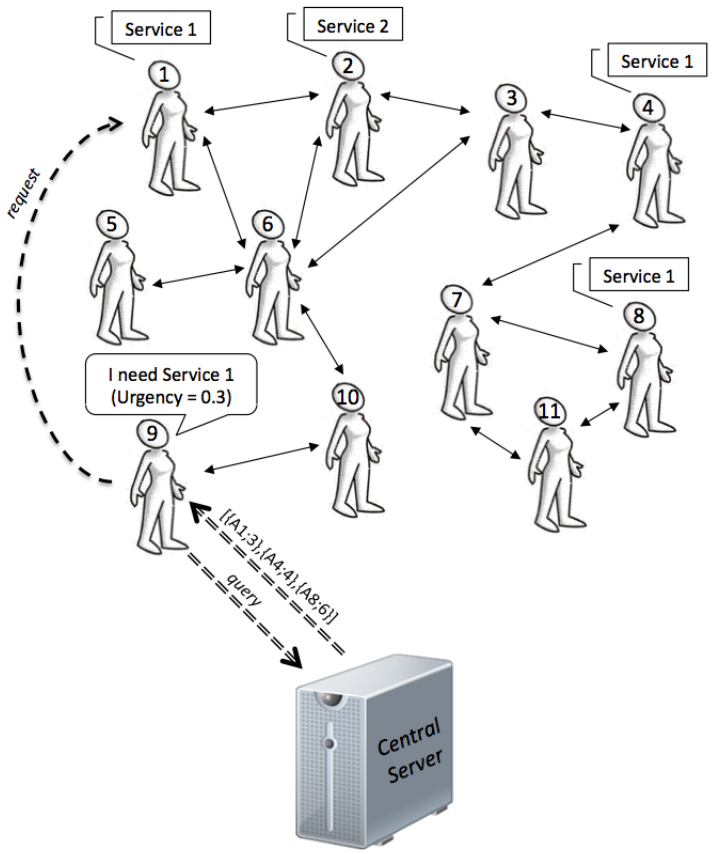


Figure 8.2: Social Network of the Web-Service Provisioning Scenario.

butions depending on the experiment. Moreover, the service's capacity and the service's expected quality follow two different probability distributions, both defined by the system and dependent on the environmental conditions that will be simulated.

Algorithm 6: Simulation Process.

```
1  for timesteps do
       /* Environmental Phase:  Resource & Needs Assignation                    */
2      forall i ∈ agents do
3          coin = random();
4          if coin ≤ serviceAssignationProbability then
5              s = createNewService(type,capacity);
6              S.add(s);
7              i.assignService(s);
8              forall j ∈ agents do
9                  if coin ≤ needAssignationProbability then
10                     j.assignNeed(s.type, deadline);
11                 end
12             end
13         end
14     end
       /* Search & Exchange Phase                                              */
15     forall i ∈ agents do
           /* Ordering the needs                                               */
16         i.orderNeedsByUrgency();
17         n = i.mostUrgentNeed();
           /* Query server about most urgent needed service                    */
18         Server ← i.serverRequest(n);
19         potentialContacts = i.receiveServerResponse() ← Server.serverResponse({c_0,d_0},{c_1,d_1},...,{c_k,d_k});
           /* Send a request to the closest of the available service holders   */
20         contact = potentialContacts.closestDistance();
21         contact ← i.sendRequest(n,n.quality,n.urgency);
           /* Receive and evaluate response for norm violations                */
22         response = i.receiveResponse() ← contact.decideRequest(i,n,urgency);
23         if response == Share then
24             i.Block(contact.distance);
25             contact.Block(contact.distance);
26         end
27         normativeResponse = i.decideNormativeReaction(response);
28     end
       /* Strategy Update Phase                                                */
29     forall i ∈ agents do
30         nc = i.ObserveNormaticeCues(i.neighbours);
31         i.UpdateStrategy(nc);
32     end
33 end
```

The simulation is run for 20.000 time steps, and each time step is structured in the following way:

1. According to the environmental situation, each agent has the possibility to provide some services and has some needs to satisfy (lines 2-14 in Algorithm 6). The needs are only assigned to agent when a service is created, and that is why we created the needs assignation after the creation of the service.

2. Each agent can ask the system for a needed service (line 18 in Alg. 6), thus receiving a list containing the service's holders (line 19 in Alg. 6). The server list has the following form $\{c_0,d_0\},\{c_1,d_1\},\ldots,\{c_k,d_k\}$, where c_i represents the agent identity, and d_i represents the distance from the requester to that agent.

3. Each agent can send to the selected service-holder a request (as it can be seen in the schematic communication protocol shown in Fig. 8.3) with a *urgency* parameter. The service-holder is chosen giving preference to those that have available services and are located at the closest distance (line 20-21 in Alg. 6).

4. To simplify our experiments, agents can take any of these three decisions: "share", "not to share" or "not allowed to share". The decision-making functions will be further discussed in Section 8.3.4.

5. If the service holder's response is positive ("share") and the quality of the received service is at least equal to the expected one, the requester becomes a service-holder, and both the original service-holder and the requester remain blocked for a fixed number of time steps (representing the transaction time); if the quality of the service is insufficient, the requester will continue to look for that service, discarding the one received (line 23-26 in Alg. 6)

6. After receiving the service holder's response, if a low quality service is received (line 27 in Alg. 6), the requester can decide to punish or sanction the service-provider. As said in section 8.3.2, punishment works by imposing a fine to the target, thus modifying the cost-to-benefit ratio of norm compliance and violation, while sanction is also accompanied with a normative message, making explicit the existence and violation of the norm to the target (and potentially to the audience).

7. At the end of each time step, agents observe the interactions that have occurred around them, this way checking the amount of norm's compliance, violation, punishment and sanction and consequently update the norm's salience mechanism (as explained in Sec. 8.2). Our agents have access to perfect local information: they can record only the social normative information of their direct neighbors in their social network (line 29-32 in Alg. 6).

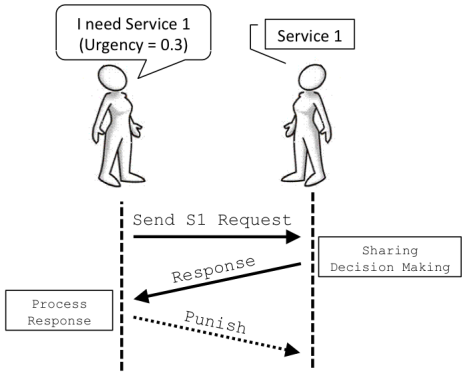


Figure 8.3: The Interaction Protocol Between Agents

The two norms that govern the environment, described in Sec. 8.3.2, are formalized as follows:

- N1, the *Always Share when Asked for a Service* norm:
  $\forall$*requesterAgent* **Share** $\leftarrow$ contact.decideRequest(requesterAgent,requestedService,urgency);

- N2, the *Share Only High-Quality Services* norm:
  $\forall$*requesterAgent* **Share** $\leftarrow$ contact.decideRequest(requesterAgent,requestedService,urgency)
  iff requestedService.capacity>requestedService.Clients.size.

8.3.4 Agents Architectures

In environments where the designer has perfect knowledge of both the system's behaviour and the agents' necessities, hardwired strategies are the best option: the system designer can use scheduling algorithms to synchronize the usage of the resources amongst the agents to obtain the optimal distribution. However, in this work, we are interested in scenarios where the dynamics of the system are unpredictable and unfold runtime. Therefore we need agents able to adapt to environmental changes. In the experimental section (Sec. 8.3.5) we compare two agents' architectures: the *Instantaneous Utility Maximizing Agents (IUMA)* and the *EMIL-I-As*.

Aiming to maximize their instantaneous utility, IUMAs share their resources only if the probability of being punished is sufficiently high. Moreover, these agents never punish (as punishing is costly and reduces their utility).

EMIL-I-As are provided with the normative architecture described in Sec. 8.2.

Before internalizing the norm, they comply with it only to avoid punishment, as IUMAs do. We refer to EMIL-I-As that are not provided the internalization modules as *normative* agents. They do not comply with norms in an automatic way, as internalizers do, they choose to comply with the norm only to avoid punishment. In other words, unlike internalizers their goal to comply with the norm is an instrumental one: it is satisfied only to avoid punishment and it is not an end in itself. Once internalized, they follow the norm also in absence of punishment. Initially, only *norm-holders* know about the norms governing their environment. Any *normative* agent (EMIL-I-A or not) can (or cannot) be a norm-holder, depending if they are initialized with knowledge about norms and the salience variable.

Therefore, *normative* agents observe norms under the threat of punishment (as this would reduce their utility). If a norm is intensively defended through the application of punishment, a normative agent will consequently observe it; on the other hand, when the norm is not defended, a normative agent will probably not observe the norm. However, agents do not know beforehand what the surveillance rates of the norm are. During the simulation, agents update (with their own direct experience and observed normative social information) the perceived probability of being punished. Normative agents' decision making is also sensible to a *risk tolerance* rate: when the perceived punishment probability is below their *risk tolerance* threshold, agents will decide to violate the

norm; otherwise they will observe it. Although this process might provide agents with the highest benefit, it yields the computational cost of evaluating at every time step each of the options. Before internalizing norms, a normative agent calculates the convenience of complying with the norm, while the salience mechanism works in the background. From the technical point of view, the normative agent will *internalize* a norm when both the following conditions are fulfilled: (1) norm salience is above a certain threshold, indicating that the norm is important enough within the social group; and (2) the cost-to-benefit calculations for all the possible actions exceed the tolerance threshold.

Once a norm is internalized, *EMIL-I-As* stop executing benefit-cost calculation, and start observing the norm in an *automatic* way. Nevertheless, the salience mechanism is still active, and it is continuously updated. This way, if necessary, agents are able to unblock the automatic action, restoring a cost-to-benefit analysis in order to decide whether to comply or not.

Internalizer Decision Making

Internalizers take two decisions: (a) whether share a service or not, and (b) whether punish or sanction norm violators. Concerning the first choice (see Alg. 7), share a service or not, if no norm has been internalized yet, the agents will evaluate all the possible actions available, and they will choose the one returning the highest payoffs (line 11-24 in Alg. 7). Otherwise, once the norm has been internalized, the agent will behave in an automatic way: the decision will be no further evaluated, but executed straightforwardly wrt the compliant behaviour of the internalized norm (line 1-10 in Alg. 7).

Concerning the second choice, when detecting a norm violation, the decision to punish/sanction[3] or not the wrongdoers depends on the level of salience of the violated norm. This decision making is described in Algorithm 8 in a very self-explanatory way.

IUMAS Decision Making

The IUMAs' decision making for the first stage decision is modelled with a classic Q-Learning algorithm (as in [Sen and Airiau, 2007, Villatoro et al., 2009]). The learning algorithm used here is a simplified version of the Q-Learning one [Watkins and Dayan, 1992].

The Q-Update function for estimating the utility of a specific action is:

$$Q^t(a) \leftarrow (1-\alpha) \times Q^{t-1}(a) + \alpha \times reward \tag{8.1}$$

where *reward* is the payoff received from the current interaction and $Q^t(a)$ is the utility estimate of action a after selecting it t times. When agents decide not to explore, they will choose the action with the highest Q value. The reward used in the learning process is the one obtained from interaction, considering also the amount of punishment received.

The IUMAs' decision making for the second stage is also modelled with a Q-Learning. However, as there are no potential risks consequent to not punishing (as in

[3]As we said, only agents having recognized that there is a norm regulating their group will sanction, otherwise they will just use punishment.

Algorithm 7: Resource Sharing Decision Making.

Input: Deciding Agent i, Requesting Agent j, Required Service s, Requested Quality q, Requested Urgency u
Data: Potential Reward for Defection $R_{Def} = 3$
Data: Potential Reward for Cooperation $R_{Coop} = 0$
`/* Automatism in case of Internalized Norms */`
if *i.Internalized(N1)* **then**
  `/* Norm 1 Compliant Behaviour */`
  Share requested resource;
end
if *i.Internalized(N2)* **then**
  `/* Norm 2 Compliant Behaviour */`
  if *s.clients.size* $<$ *s.capacity* **then**
    Share requested resource;
  else
    Do not Share;
  end
end
`/* If no norm has been internalized, Benefit-Cost Calculation considering the presence of Norms */`
`/* Agents have to update their Beliefs about Punishment Policies */`
i.updateBeingPunishedProb();
if *i.RiskTolerance* $\leq$ *i.perceivedPunishmentProb(N1)* $\vee$
i.RiskTolerance $\leq$ *i.perceivedPunishmentProb(N2)* **then**
  $R_{Def} \leftarrow$ Deduce Punishment Cost;
end
if $R_{Def} \geq R_{Coop}$ **then**
  Do not Share;
else
  if *i.Salience(N2)* $\geq$ *i.Salience(N1)* $\wedge$ *i.Salience(N2)* $\geq$ *Norm Activation Value* **then**
    Not allowed to Share;
  else
    Share;
  end
end

Algorithm 8: Decision Making in front of norm violations

Input: Requesting Agent i, Requested Agent j, Required Service s, Requested Quality q_{req}, Requested Urgency u, response, Obtained Quality q_{rec}

```
   /* Response against an Excused Service Denial                          */
 1 if response == Not allowed to Share then
       /* If low quality was requested...                                 */
 2     if q_req < 1 then
           /* And N1 is more important, it is considered as a Punishable
              Violation                                                    */
 3         if i.salience(N1) > i.salience(N2) then
 4             if i.salience(N1) ≥ ACTIVATION VALUE then
 5                 i.sanction(j);
 6             else
 7                 i.punish(j);
 8             end
 9         end
10     end
11 end
   /* Response against a Service Denial                                   */
12 if response == Do Not Share then
       /* If low quality was requested...                                 */
13     if q_req < 1 then
           /* And any norm is active, it is considered as a Punishable
              Violation                                                    */
14         if i.salience(N1) ≥ ACTIVATION VALUE then
15             i.sanction(j);
16         else
17             i.punish(j);
18         end
19     else
20         if i.salience(N2) ≥ ACTIVATION VALUE then
21             i.sanction(j);
22         else
23             i.punish(j);
24         end
25     end
26 end
   /* Response against a Low Quality Sharing                              */
27 if response == Share ∧ q_rec < q_req then
       /* If it is not an urgency...                                      */
28     if u¡i.salience(N2) then
           /* And N2 is active, it is considered as a Punishable Violation */
29         if i.salience(N2) ≥ ACTIVATION VALUE then
30             i.sanction(j);
31         else
32             i.punish(j);
33         end
34     end
35 end
   /* If no normative action has been taken, and any norm is highly salient,
      educational messages are sent                                       */
36 if No Punishment/Sanction Yet then
37     if i.salience(N1) > i.salience(N2) ∧ i.salience(N1) ≥ ACTIVATION VALUE then
38         i.educate(j);
39     end
40     if i.salience(N2) > i.salience(N1) ∧ i.salience(N2) ≥ ACTIVATION VALUE then
41         i.educate(j);
42     end
43 end
```

the meta-punishment situation presented by [Axelrod, 1986]), agents will always prefer not to punish.

8.3.5 Experimental Design

In the simulation platform presented, two important factors can vary, thus affecting the system's behaviour: (1) the services' capacity, (2) the expected desired-quality of the services.

In order to check the adaptability of EMIL-I-As, the service's capacities and its expected desired-quality vary during the simulations.

To reduce the search space, the service and task assignment have been fixed to a constant rate of 10%: the resources are created at an average rate of 1 every 10 time steps, and at that time step, 10% of the population will be assigned with tasks that need that resource. Tasks are assigned only when resources are created.

Agents are located in a scale-free network (that represents theoretical social networks [Newman, 2003, Albert and Barabasi, 2002]). This topology restricts the agents observation (with respect to the social and normative information) to their direct neighbors, but they can potentially interact (ask and receive services) with any other agent in the network.

In this work, the cost of punishment has been fixed to 1:4, meaning that punishing costs 1 unit to the punisher while reducing the target endowment of 4 units (the 1:4 punishment technology is used because found [Nikiforakis and Normann, 2008] more effective in promoting cooperation). The decision making of both normative agents and internalizers is not affected by the cost of applying a punishment/sanction to the target; however, the costs of being punished/sanctioned affects the agents' decision to cooperate or defect in the next rounds. The results presented in this section are the average results of 25 simulations.

All non-IUMAs agents are initialized with a constant propensity to violate norms and the perceived probability of being punished/sanctioned is equal to or lower than 30%. From the beginning of the simulation, some agents (this quantity varies from experiment to experiment) are already endowed with the two norms, N1 *Always Share when asked for a Service* and N2 *Share Only High-Quality Services*. We refer to these agents as norms' holders and the initial salience of their norms is set to 0.9. Before proceeding and for the sake of clarification, as specified in Sec. 8.3.4, a normative agent is an agent endowed with the EMIL-I-A architecture and that has not internalized the norm yet, while with internalizers we refer to normative agents that have the internalization module and can reach the internalization state and comply with norms in an automatic way. Both types of agents (internalizers and normative) can (or cannot) be norm-holders, i.e. knowing the existence of the norm with a salience value assigned to it.

The simulations are populated with a constant amount of 50 agents, with a variable amount of internalizers (whether or not norm holders) and IUMAs.

8.3.6 Experiment 1: Adapting to Environmental Conditions

One key feature of our internalizers is their ability to dynamically adapt to unpredictable changes. To test this ability, we have designed several dynamic situations, where the providers' capacities and the requesters' desires vary, in order to dynamically change the norm that apply along the simulation without the knowledge of the agents. From the beginning of the simulation, the environment is fully populated by internalizers, already having the two norms highly salient stored in their minds (norms holders) [4].

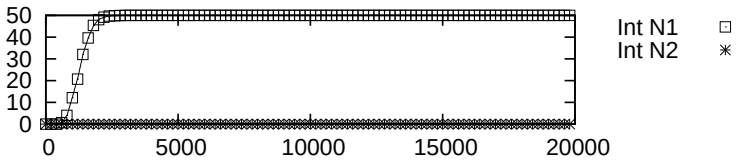
In Fig. 8.4, on the x-axis the timesteps of the simulation are shown, on the y-axis the amount of internalizers is indicated.

Firstly, in Fig. 8.4(a) we show the dynamics of a basic scenario, where the services are endowed with *infinite* capacity to attend clients, and the requesters do not care about the quality of the services they receive (meaning that also services with a low quality satisfy their needs). We can observe that, after a number of timesteps the internalization process starts working, resulting in an increased number of agents that internalize norm 1 (Always share when asked for a service). Why is norm 1 spreading and being internalized rather than norm 2 (Share only high-quality services)? This is due to the fact that in this experimental condition agents looking for a service are not interested in its quality, thus even though they receive services with a low quality their needs are satisfied and they will not punish/sanction the service holder that provided it. Sharing a service, whichever its value is, is interpreted as an action compliant with norm 1 and since instances of this action exceed the number of actions in which high-quality services are shared, the salience of this norm is higher than the other's. This experiment confirms that our internalization architecture works selecting the correct norm to internalize.

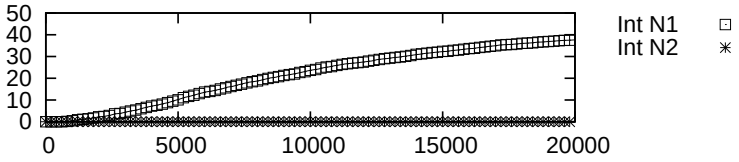
The second experiment is a slight but important variation of the previous one. This time, the services' capacity is restricted (to 3 clients). Results in Fig. 8.4(b) show that agents correctly internalize the norm that rules the situation (i.e. norm 1), although the internalization process is slower than in the previous experiment. The reason for delay lays in that the dynamics taking place in this experiment: no clear information is given to the agents for deciding which one, between norm 1 and norm 2, is governing the system. Internalizers are initialized with both norms at high salience. When the resources' holders start sharing, they provide requesters with high quality resources, even though the requesters' needs would be satisfied also by low quality resources. Agents interpret high-quality transactions as actions compliant with norm 2, this increasing its salience value. After a while, agents start sharing resources of lower quality, because the list of requesters exceeds the amount of high quality resources. Low-quality resource transactions make the salience level of norm 2 decrease and the salience level of norm 1 increases to its maximum value (necessary condition for norm internalization as shown in Sec.8.2.1). We can observe how all these micro-dynamics in the individuals affect the global behaviour of the system producing the delay showed in Fig. 8.4(b).

In the third experiment, we test the agents' capacity to internalize norm 2: in this situation the services' capacity is restricted to 3 clients and the requesters' needs are satisfied only when receiving high-quality resources. In general, we observe that the

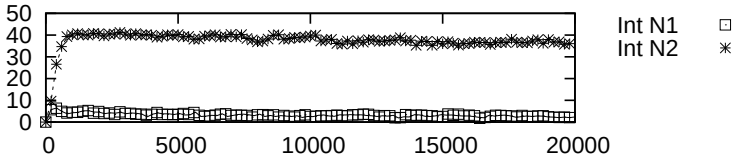
[4] As we will see in Sec. 8.3.8, the internalization process speed is proportional to the initial amount of norm holders.



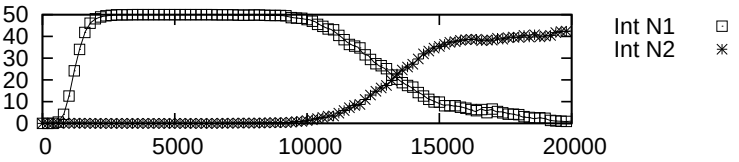
(a) **Experiment 1**: Infinite services' capacity and expected low quality of the services (meaning that agents need to have the service without paying attention to its quality)



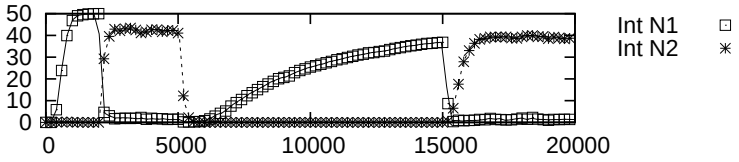
(b) **Experiment 2**: Restricted services' capacity (restricted to 3 clients) and expected low quality of the services



(c) **Experiment 3**: Restricted services' capacity (restricted to 3 clients) and expected high quality of the services



(d) **Experiment 4**: Services's capacity is *variable* along the simulation: the services' capacity is infinite during the first 2000 timesteps, and after that moment, it linearly decreases from 20 to 1, reaching the service exploitation point (where they cannot longer offer high quality services) after timestep 8000. Agents are assigned tasks to be fulfilled at *high* quality level along the entire simulation.



(e) **Experiment 5 (The Complex Loop)**: Before the timestep 2000, the capacity of the services is infinite, but it is set to 3 after that. Before timestep 5000 and after the timestep 15000, agents are assigned needs that are satisfied also with low quality services. During the rest of the execution, the expected quality of the services is high.

Figure 8.4: Different Combinations of Resources Capacities Distribution and Expected Quality Distributions.

agents internalizing norm 2 do not reach the majority, but when this is the case they do so quickly. This is because until the number of resources is high enough to satisfy all the requests with their expected quality, being compliant with norm 2 can be interpreted also as an action compliant with norm 1, thus maintaining the salience of norm 1 high.

In the fourth experiment a dynamic change in the environment is included. The system is programmed in such a way that for the first 2000 timesteps the capacity to provide resources is infinite; after that time, during the simulation, the capacity linearly decreases from 20 to 1 until. Agents' needs are satisfied only when they receive high-quality resources. We observe that after timestep 8000, the capacity of the services significantly decreases and agents slowly start deinternalizing norm 1 and substitute it with norm 2. This experiment shows how our agents are able to adapt to this dynamic situations.

In order to test their speed of adaptability, we have designed one last experiment. In Fig. 8.4(e), we show the result of the experiment named *The Complex Loop*, where we can observe that our internalizers perform efficiently even in complex situations, i.e. where not only the environment (the services capacity), but also the agents' preferences (desired-quality of the services) change during the simulation.

The *Complex Loop* experiment shows that when the services' capacity is infinite (meaning that high-quality services are always available) and the agents' needs are satisfied by high-quality (HQ) resources, *norm 1* is internalized. At timestep 2000, an abrupt change occurs in the environment, making the services' capacity drop to 3, and agents switch to internalize *norm 2*. After time step 5000, agents' preferences change (starting now to be satisfied also with Low-Quality resources), producing another change in the internalization dynamics. Agents turn back to internalize *norm 1*. Finally, agents' preferences change again at time step 15000, preferring HQ resources and internalizing *norm 2* again.

We can conclude that a population of internalizers can adapt to sudden and unexpected environmental changes in a flexible manner.

8.3.7 Experiment 3: Internalizers vs IUMAs

In this experiment, the internalizers' performances are compared with the IUMAs ones. We run several simulations to cover the search space of different proportions of Internalizers[5] and IUMAs.

As it is shown in Fig. 8.5, a higher amount of Internalizers is able to adapt faster to the changing environmental conditions. The figure shows the percentage of unsuccessful transactions[6] occurred in the system during a situation identical to the one presented in the "Complex Loop" experiment, except that the population is a combination of Internalizers and IUMAs.

The results show that the number of unsuccessful transactions is inverse-proportional to the number of internalizers. The explanation of this result can be found in the way in which IUMAs decide to comply or not with the norm: when they decide to

[5]Also in this scenario, Internalizers are also norm-holders

[6]We consider transactions as unsuccessful when a service provider offers a service of a quality lower than the expected one, or when a service, though available, is not shared .

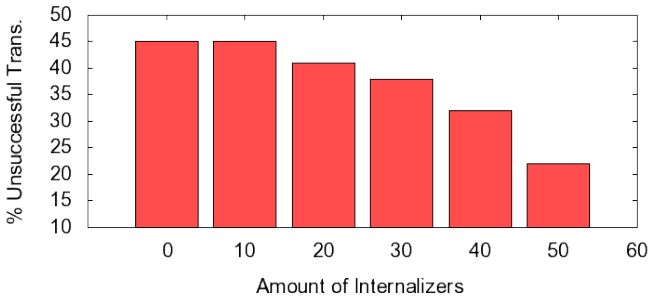


Figure 8.5: Percentage of Unsuccessful Transactions with Different Proportions of Strategic Agents and Internalizers.

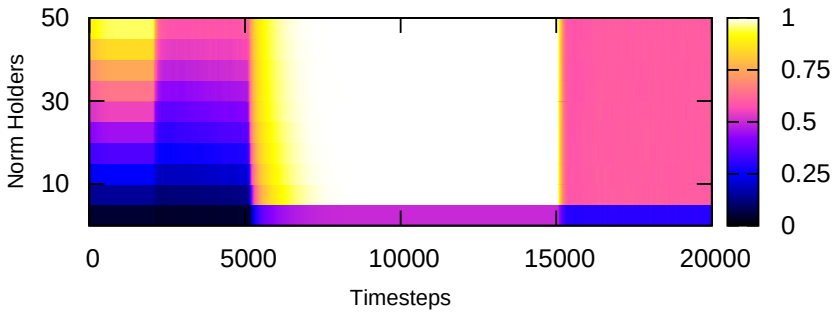
share a resource their payoff is reduced (when sharing, agents are blocked for a number of timesteps), therefore they decide to free-ride. In other words, they always ask for services, but they never share.

On the other hand, internalizers decide to comply or violate the norm, not only to maximize their material payoffs, but also because they recognized a norm in their social environment, and form the goal to comply with it, depending on its salience. Consequently, when the environment change and agents need to find out the new norm, internalizers perform a certain number of unsuccessful transactions before adapting to the new situation. The simulation data show that the unsuccessful transactions are 50% less (from 45% to 22,9%) in groups entirely populated by internalizers than in those populated by IUMAs.

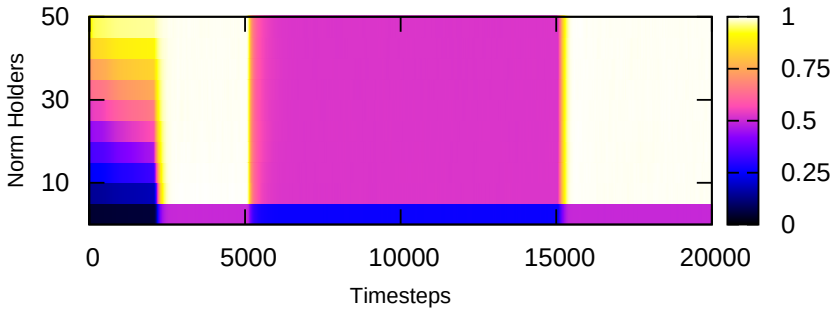
8.3.8 Experiment 3: Effect of Initial Norm Holders

This experiment aims to analyze how the initial amount of norm holders affects the performance of a group fully populated by internalizers. As explained in Sec. 8.2, internalizers do not know which are the norms in force in the group at the simulation onset, therefore the presence of norm-holders, explicitly eliciting the norms, is necessary for triggering the process of norm recognition.

Focusing again on the *The Complex Loop* experiment (presented in Sec. 8.3.6 and shown in Fig. 8.4(e)), in Fig. 8.6 the average norm's salience per agent is shown: on the x-axis the simulation time evolution is shown, while on the y-axis the amount of initial norm holders (randomly located in the network) is indicated. We obtained similar results in the rest of experiments performed in Sec. 8.3.6, but we decided to analyze only this situation because of its complexity and completeness (as it contains all the possible dynamics of our scenario with respect to norm internalization and de-internalization). Figure 8.6(a) represents the salience dynamics of norm 1 and Figure 8.6(b) that of norm 2: depending on the moment of the experiment, only one of the two norms is the one that really prescribes how to behave; that is why its salience is higher than the other one (for a more detailed description of the relation between the two norms, see, Sec. 8.3.6).



(a) Norm 1 Average Salience



(b) Norm 2 Average Salience

Figure 8.6: Complex Loop Experiment Average Salience.

Figure 8.7: Different norms coexisting in the same social environment.

During the first 2000 timesteps of the simulation, we can appreciate the impact of a different number of norms' holders on norms' salience: when this number increases, so does the norm's salience. Special attention should be paid to the situation with no initial norm-holders: here norms are never recognized as no explicit norm elicitation occurs (by explicit norm invocation or by sanction). Even a small number of initial norm-holders (10% of the population in this case) allows agents to recognize the norms and internalize them. Once recognized and stored according to their salience degree, agents will start both comply with the norms and also defend them. Thus, norm holders are necessary to trigger a virtuous circle from compliance with the norms to their enforcement. After the two norms have been recognized (around timestep 5000), the number of initial norm holders will no more affect the successive dynamics.

8.3.9 Experiment 4: Testing Locality: Norms Coexistence

In self-organized societies, as are the ones we are interested in, norms are an important tool for solving the dilemma of public good provision. Since individual behaviours affect the group's welfare, the group itself needs to exercise control over its members. Consequently, agents can create and impose a norm depending on the goals of the society. Moreover, in this type of distributed systems, different dilemmas could appear in different parts of the social network, depending (again) on the distribution of population's preferences distribution and the environmental conditions. Our internalizers are able to cope with the locality of the norms, thus allowing the coexistence of different competitive norms in the same social network. To observe this result we have performed the following experiment.

We have placed 60 agents in a one-dimensional lattice, with neighborhood size set to 6 (i.e. each agent has a constant number of 6 neighbors). In Fig.8.7, agents in the top right and in the bottom left quarters of the ring are assigned with infinite capacity resources and low quality desired services, and, the rest of the agents are given limited capacity services and high-quality desired tasks. The colour of the nodes indicates the internalized norms (light colour standing for norm 1, and dark colour for norm 2). The

self-adaptive capability of our internalizers shows a good performance in the designed environment: two norms can coexist in different areas of the same social network.

8.3.10 Experiment 5: Dealing with Emergencies: Selective Norm Violation

One of our claims is that internalizers are able to unblock a norm not only in situations where its salience is very low, this meaning that the norm is disappearing, but also in emergency situations.

While IUMAs cannot handle normative requests with different levels of urgency (because they are not endowed with normative architectures), our internalizers (thanks to the salience module) can. In the ambulance case, the urgency of the situation is announced by the siren, thus allowing agents to unblock the automatic behaviour of stopping at the red traffic light (and the ambulance to pass), as the emergence situation is more salient than the observance of the norm. In our distributed web-service scenario, we can imagine the following situation: imagine one of our agents is writing a scientific paper which includes a number of important calculations. However, the day before the deadline the agent in question realizes it needs to rerun some of the experiments; in order to get the results in time, it needs to use different calculation clusters distributed amongst its peers, who are requested accordingly. The requested agents can choose to follow the FIFO (first-in, first-out) norm, or, given the urgent situation, make an exception and proceed to the execution of our last-minute scientist's calculations. It is easy to see how this norm violation performed by its neighbours/colleagues would bring a benefit to our agent (as it will have the results in time for submission).

Experimental results show us that the internalizers successfully handle emergencies when these are explicitly specified (as the siren in the ambulance case). However, there is a cost associated to this advantage. An emergency is interpreted by the audience as a non-punished norm-violation, consequently reducing the norm' salience (stopping at the red traffic light). If agents face the same emergency too often, the salience of the norm "normally" regulating the situation will decrease and this reduction unblock the internalization process, leading the agent back to the normative (benefit-to-cost calculation) phase. We can then affirm that emergencies can be managed by internalizers, but only when they occur sporadically, otherwise they are interpreted as a change in the salience of the norm governing the environment.

8.3.11 Experiment 6: Topological Location

Finally, we pay attention to the importance of the location of our INHs (initial norm holders) in the topology, when dealing with scale-free networks. As agents handle local information to adapt the notion of norm activation, those agents located closer to INHs will have a stronger and faster norm recognition process and salience adaptation. In other chapters of this thesis (Chapter 4), we identified that agents located on certain positions within a Scale-Free network played an important role on the emergence of conventions. We hypothesize that in irregular networks, such as scale-free, agents with higher connections will be the most adequate to be initialize as INHs, as they will be

able to influence a larger portion of the population, producing a faster spread of norms within the population.

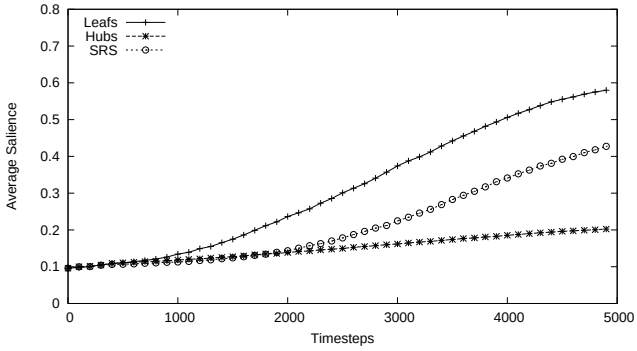
In order to analyze the effect of the topological location of INHs, we adjust our experimental platform in order to position our INHs in certain locations of the topology and contrast their effects:

- *Hubs*: INHs will be located in the nodes with the highest degree.
- *Leafs*: INHs will be located in the nodes with the lowest degree.
- *Self-Reinforcing Structures (SRSs)*: INHs will be located in the nodes with a higher betweeness inside a Self-Reinforcing Structure. Eventhough the concept of Self-Reinforcing Structure have sense in the convention emergence problem, we believe that the position of these agents (connecting small communities with the rest of the society) can have an important impact on the distribution of norms.

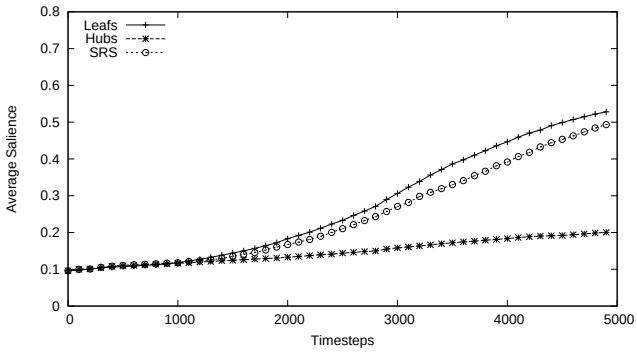
In order to obtain the results, we run each of the treatments 25 times and plot the average results. As we want to observe the effect of the location of the INHs in the distribution of norms, we load a society fully populated with internalizers, and we vary the number and position of INHs. Agents are located in the first scenario used in Sec.8.3.6, where services capacities are infinite and it is expected low quality from agents; that way we only pay attention to the average salience of *Norm 1: Always Share when Asked for a Service*.

Experimental results shown in Figure 8.8 are against our initial hypothesis, showing that locating INHs on the leafs produce a stronger effect on the average population's salience. Even though it seems counterintuitive, the explanation is straightforward. Because of the design of our agents, INHs decisions are motivated by the information gathered at a local level. In the *Hubs* treatment, INHs access more information than in the *Leafs* treatment, as the INHs have more neighbours in the former treatment. For that reason, hub INHs initially observe more violations than the others. On the other hand, leaf INHs are only connected to very few agents, and therefore the information that they gather is reduced with respect to the Hub INH. Moreover, and using an example, if a leaf INH is connected to only one agent, and this one is a violator, the INH will punish/sanction this violation. As we saw previously, a punished/sanctioned violation have a strong impact on salience. On the other hand, the hub INH that has a large number of neighbours will observe a reduced number of punished/sanctioned violations, remaining the other unpunished. As we have said, those actions have an effect on the salience of INH that motivate their decisions. In the case of Hub INHs, INHs loose their motivation to punish/sanction faster than a leaf INH.

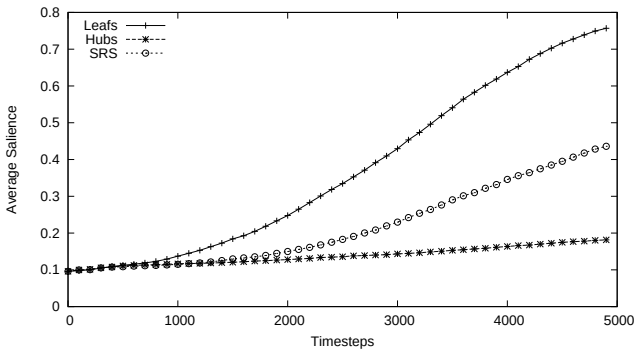
These results suggest us that using punishment in a self-organizing society should be a bottom-up process, so that agents do not loose the motivation to punish/sanction free-riders. Moreover, this result is encouraging for policy-makers, as it is not necessary to locate INHs in hub positions of a network (which are hard to obtain because of the degree of these nodes), being the correct position the leafs of the network.



(a) 10 INHs.



(b) 30 INHs.



(c) 50 INHs.

Figure 8.8: Effect of the topological positioning of INHs.

8.4 Conclusions

When Vygotsky first formulated his theory of internalization, he noted that only *"the barest outline of this process is known"* [Vygotskii and Cole, 1978, p 57]. We do not know, yet, how people manage to internalize beliefs and precepts with a reasonably adequate success, partly because we still do not agree about what to investigate, or what we mean by this notion. Consequently, no useful notion and model are available for applications, despite its wide and profound implications. Questions such as how norm internalization unfolds, which factors elicit it, which are its effects, obstacles and counter-indications, are issues of concern for all of the behavioral sciences. The internalization of social inputs is indispensable for the study and management of a broad spectrum of phenomena, from the development of a robust moral autonomy to the investigation and enforcement of distributed social control; from the solution to the puzzle of cooperation, to fostering security and fighting criminality, etc.

After a cognitive definition of the subject matter, this chapter presented and discussed the building blocks of a rich cognitive model of internalization as a multi-step process, including several types and degrees of internalization. Next, factors favoring different types of internalization were discussed. The modular character of BDI architectures, like EMIL-A was shown to fit the approach advocated in the paper.

In this work we have also implemented and presented the results of the internalization module that has been incorporated into the existing platform, creating the new EMIL-I-A (EMIL Internalizer Agent). Internalizers have been tested in a tragedy of the digital commons' scenario, where the emergence of cooperation is a difficult task to achieve and they have proven to successfully deal with this task. The simulation data have also shown that when facing situations in which the environment can quickly change internalizers are able to self-adapt. We have also observed that the amount of initial norm holders and internalizers speed up the process of norm's convergence, even when the scenarios are dynamic and different norms have to be internalized and de-internalized. Another interesting result is how the subjective character of the salience allows the coexistence of different norms in the same social environment: depending on the agents' interests and the environmental conditions, different norms can emerge in the same environment. Finally, the experiments have shown that our internalization architecture is flexible enough to handle emergencies and decide to violate an already internalized norm. We claim that internalization provides agents with self-policing capabilities that are very useful in settings where social control is unfeasible and expensive.

What is the added value of a rich cognitive model of internalization, as compared to simpler ones (e.g., reinforcement learning)? There are several competitive advantages. First, reinforcement learning does not account for the main intuition shared by different authors, i.e. the idea that internalization makes compliance independent of external enforcement. Second, a rich cognitive model, namely a BDI architecture with its specific modules, accounts for different types and degrees of internalization, bridging the gap between self-enforcement and automatic responses. Third, a BDI architecture accounting for different levels of internalization allows flexibility to be combined with automatism, as well as thoughtless conformity with autonomy. A BDI system can host automatism, but a simpler agent architecture does not allow for flexible, innovative and

autonomous action.

Chapter 9

Conclusions

In this final chapter we summarize the main contributions of the thesis that have already appeared partially in the respective sections of the dissertation. We also state the future lines of research that this thesis has opened. For the sake of clarity we will divide both the conclusions and the future works in two sections, one for the convention emergence part and the other for the emergence of cooperation in Multi-agent Systems. However, there is a main conclusion to be extracted from the thesis as a whole: we have presented a game theoretical differentiation for social norms, and a different range of techniques that can be used by agents (independently) in order to achieve their tasks in a more efficient manner. This thesis shows how norms are essential for the achievement of coordination between agents without a centralized authority, even though these norms are against the agents' self interest.

9.1 Convention Emergence

9.1.1 Conclusions

In this thesis we have developed a general framework for the convention emergence problem. The problem of convention emergence is a key problem to be solved for real self-organized systems. By allowing agents to reach a common convention from all possibles is one of the simplest coordination tasks to be achieved in a decentralized way. In our approach we focus on the emergence of the conventions following a social learning approach, as it seems to be the most realistic and less intrusive for the rest of agents in the population (others have studied this problem although accessing to private information of the other agents). Our experimental framework has been designed to vary a number of parameters that were never previously explored in the literature (type of game, players per interaction, conventions available, strategy update rules, learning algorithms and topologies), and we believed were essential for a complete understanding of the phenomenon.

The systematic variation of these parameters allowed us to analyze the dynamics of the process of convention emergence and analyze the reasons why previous researchers

did not overcome the threshold of 90% emergence rate. We have seen how certain configurations of the social network and some strategy update rules promote the emergence of subconventions that remain meta-stable. It is the first time during the our research in convention emergence that we identify this meta-stability, introduced by the social learning and the topology of interactions.

In order to dissolve those subconventions we analyzed the state-of-the-art of the socially-inspired instruments used in MAS and propose two new ones: rewiring and observance. Rewiring allows agents to break links with agents from which they are not getting satisfactory interactions, while Observance allows agents to obtain information about the last played strategy of a certain subset of agents. With the usage of these instruments independently, the emergence of global conventions is accelerated in some scenarios, remaining others meta-stable. Specifically, we observe than when our instruments are only fed with local information, the subconventions remain meta-stable.

The recognition of this locality effect took us to discover the existence of Self-Reinforcing Structures in Scale-Free Networks. These structures are a group of agents that can emerge a subconvention that remains stable because of the interaction topology which directly affect to the social learning. The Self-Reinforcing Structures detected are *Claws* and *Caterpillars* in Scale-Free Networks. With an instrument that combines the power of observance and rewiring, we achieved the dissolution of the Self-Reinforcing Structures, allowing then the system to achieve full-convergence, accessing only public and local information.

9.1.2 Future Work

Along this work, we have seen how the convention emergence problem show interesting dynamics when studied in complex networks such as Scale Free. As this type of networks represent theoretical social networks, the future work regarding this area of research involves a careful analysis of the convention emergence in Scale-Free Networks.

In this thesis we have already thrown some light about the dynamics observed in those type of networks, and how the existence of Self-Reinforcing Structures delayed the process of convention emergence. Our solution (social instruments) proposes the dynamic rewiring of the network. However, we did not concentrated on the maintenance of the characteristics that define such networks: scale-free networks are characterized by a number of properties (short average network diameter, small clustering). Our rewiring instrument does not pay attention in maintaining those properties which determine a scale-free network. Recent work [Scholtes, 2010] presented a distributed rewiring scheme that is suitable to effectuate scale-free overlay topologies. One of our short-term tasks is the development of an algorithm that produces effective rewiring, maintaining the scale-free network, accessing only local information.

Another line of research that we consider of vital importance to be performed is the analysis of reconfiguration techniques. This thesis has studied how conventions are achieved under static environments. It would be interesting to study how our social instruments ease the process of reconfiguration to different conventions if one of them looses its validity within the social context of agents.

Finally, we plan to introduce more complex protocols to be learnt by agents: in this part of our work, agents only had to learn how to behave in a single action interaction. However, we plan to introduce norms that involve more than one action, allowing agents to learn normative patterns with the form of protocols. By studying the complexity of the protocols (the number of available actions, the length of the protocol and the typology of agents) we will observe how our agents are able to coordinate for more complicated tasks.

9.2 Essential Norms and Emergence of Cooperation

9.2.1 Conclusions

This thesis have shown a game-theoretical approach to norms that differentiates between conventions and essential norms, as games that represent different dynamics with respect to the interest of the individuals. We have seen how conventions are not a solution for certain phenomena, like situations that represent common-good games. Initially we envisioned an specific type of peer punishment (distributed punishment) to impose the emergence of cooperation and its maintenance. Contrary to monolateral punishment technologies (like that used in [Dreber et al., 2008]), distributed punishment seemed attractive since it would reduce individual costs (by dividing the cost of punishment amongst the punishers), obtaining higher social costs (by making defection a less attractive option for norm-violators).

The empirical results obtained with human subjects showed us that motivations towards cooperation are not only driven by the external enforcement, but affected by other factors. Our hypothesis was that certain types of peer punishment have a stronger impact on the agents decision making than the utilitarian damage.

In this thesis we have established the distinction between punishment (utility deduction) and sanction (utility deduction plus norm elicitation), and we have sketched a BDI agent architecture (EMIL-I-A) whose goals (and beliefs) are not only affected by a benefit-cost calculation, but also, by the existence of norms (and their importance within the social context).

Experiments with human subjects confirmed the hypothesis of the difference between punishment and sanction, using a experimental platform where humans could participate in the experiments through an electronic institution. These experiments have also served us to fine-tune the parameters that regulate the agent architecture, proving another successful example of cross-fertilization between experimental economics and computer science [Grossklags, 2007]. This differentiation between punishment and sanction is an important step forward, specially in the economics research, as so far little importance was given to the non-costly communication messages (assuming them as *cheap talk* [Farrell and Rabin, 1996]). The usage of sanction produces a more resilient individual, as its decision making is also influenced by normative motivations.

We exploit the capabilities of EMIL-I-A in simulated scenarios, proving that our agents achieve the emergence of cooperation with lower social costs and in a more resilient way. We have performed a detailed analysis varying the amount of our agents in populations of pure utilitarian agents, observing the adaptive capabilities of our ar-

chitecture. Another interesting contribution of this work is the intelligent Dynamic Adaptation Heuristic which allows agents to dynamically change the amount of punishment/sanction sent to norm-violators, producing an efficient punishment mechanism. One final interesting result confirm the robustness of sanction against invasions of free riders into the population.

Finally, we have studied how *Norm Internalization* can be incorporated as a different module in the agents mind allowing them to avoid a continuous evaluation of the possible actions in their decision making process. This internalization mechanism allows agents to save computational resources by firing automated actions in certain *internalized* situations. We have studied how internalizers perform in different types of scenarios, achieving good performance without sacrificing adaption skills. Moreover, the final experiments have allowed us to test the locality of social norms, being possible the emergence of different norms in different areas of the topology.

9.2.2 Future Work

The emergence and maintenance of cooperation is a problem that worry policy makers of open distributed punishment, as the natural tendency in these type of systems is to defect and free-ride, producing a collapse of the system. As an example, when Gnutella was one of the most popular p2p sharing platform, 70% of users did not share resources [Adar and Huberman, 2000]. In this thesis we have studied how different punishment technologies can affect the decision making of agents in a distributed manner, without the necessity of a centralized authority, which would be costly for the system. Unfortunately, the sanctioning technology that we propose also implies a cost for the punisher, entering in a second-order public good game. For that reason, and as a future work to be performed in this area of research, we are interested in studying how reputation can work as a cost-free mechanism that ensures cooperation. By providing agents with the ability to communicate, they can send evaluations that affect their future interactions with other agents. Agents need to be given with the trust and reputation mechanism that (1) aggregates the rumors that receives, and (2) activates a broadcasting mechanism to spread rumors. By giving agents with this mechanism, agents also need to be given an heuristic to determine what to communicate and to whom. This problem is essential to be understood within the scope of social networks.

As a more direct task in our future work plan we need to carefully analyze the Dynamic Adaptation Heuristic used for the efficient sanctioning. At the moment, the heuristic changes the damage associated to punishment and sanction depending on the amount of norm-violators within their social environment. However, we believe that this adaptation should be also proportional to the changes in the environment in order to be more efficient: if a larger amount of norm-violator appears, the damage of sanction should be larger (instead of increasing gradually, producing a number of punishment acts with no deterrent effect).

Bibliography

[nap, 2006] (2006). Napster. http://www.napster.com/.

[Nor, 2008] (2008). Special issue on normative multiagent systems. In van der Torre, L., Boella, G., and Verhagen, H., editors, *Journal of Autonomous Agents and Multi-Agent Systems*, volume 17. Springer.

[Adar and Huberman, 2000] Adar, E. and Huberman, B. A. (2000). Free riding on gnutella. First Monday.

[Albert and Barabasi, 2002] Albert, R. and Barabasi, A.-L. (2002). Statistical mechanics of complex networks. *Review of Modern Physics*, 74(1):47–97.

[Albert et al., 2000] Albert, R., Jeong, H., and Barabasi, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382.

[Aldewereld et al., 2007] Aldewereld, H., Garcia-Camino, A., Dignum, F., Noriega, P., Rodriguez-Aguilar, J. A., and Sierra, C. (2007). Operationalisation of norms for usage in electronic institutions. In *Coordination, Organization, Institutions and Norms in agent systems, Lecture Notes in Computer Science*.

[Aldewereld et al., 2005] Aldewereld, H., Grossi, D., Vázquez-Salceda, J., and Dignum, F. (2005). Designing normative behaviour by the use of landmarks. In *Proceedings of AAMAS-05 International Workshop on Agents, Norms and Institution for Regulated Multi Agent Systems*, pages 5–18.

[Andrighetto et al., 2010a] Andrighetto, G., Campennì, M., Cecconi, F., and Conte, R. (2010a). The complex loop of norm emergence: A simulation model. In Deguchi, H. and et al., editors, *Simulating Interacting Agents and Social Phenomena*, volume 7 of *Agent-Based Social Systems*, pages 19–35. Springer Japan.

[Andrighetto et al., 2007] Andrighetto, G., Campennì, M., Conte, R., and Paolucci, M. (2007). On the immergence of norms: a normative agent architecture. In *Emergent Agents and Socialities: Social and Organizational Aspects of Intelligence. Papers from the AAAI Fall Symposium*.

[Andrighetto et al., 2010b] Andrighetto, G., Villatoro, D., and Conte, R. (2010b). Norm internalization in artificial societies. *AI Communications. (In press)*.

[Androutsellis-Theotokis and Spinellis, 2004] Androutsellis-Theotokis, S. and Spinellis, D. (2004). A survey of peer-to-peer content distribution technologies. *ACM Comput. Surv.*, 36:335–371.

[Araujo and Lamb, 2008] Araujo, R. and Lamb, L. (2008). Memetic networks: Analyzing the effects of network propoerties in multi-agent performance. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*.

[Aronfreed, 1968] Aronfreed, J. M. (1968). *Conduct and conscience; the socialization of internalized control over behavior [by] Justin Aronfreed*. Academic Press, New York,.

[Asch, 1955] Asch, S. (1955). Opinions and social pressure. *Scientific American*, 193(5):31–35.

[Ashlock et al., 1996] Ashlock, D., Smucker, M. D., Stanley, E. A., and Tesfatsion, L. (1996). Preferential partner selection in an evolutionary study of prisoner's dilemma. *Biosystems*, 37(1-2):99 – 125.

[Axelrod, 1981] Axelrod, R. (1981). The emergence of cooperation among egoists. *The American Political Science Review*, 75(2):pp. 306–318.

[Axelrod, 1986] Axelrod, R. (1986). An evolutionary approach to norms. *The American Political Science Review*, 4(80):1095–1111.

[Babaoglu and Jelasity, 2008] Babaoglu, O. and Jelasity, M. (2008). Self-* properties through gossiping. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 366(1881):3747–3757.

[Bandura, 1976] Bandura, A. (1976). *Social Learning Theory*. Prentice Hall.

[Bandura, 1991] Bandura, A. (1991). Social cognitive theory of moral thought and action. In Kurtines, W. M. and Gewirtz, J. L., editors, *Handbook of moral behavior and development*. Lawrence Erlbauml, Hillsdale, NJ.

[Banerjee et al., 2005] Banerjee, D., Saha, S., Sen, S., and Dasgupta, P. (2005). Reciprocal resource sharing in p2p environments. *Proceedings of AAMAS'05. Utrecht. Netherlands*.

[Barabasi and Bonabeau, 2003] Barabasi, A. and Bonabeau, E. (2003). Scale-free networks. *Scientific American*, 288(5):60–9.

[Barabási and Albert, 1999] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.

[Barkow et al., 1995] Barkow, J. H., Cosmides, L., and Tooby, J., editors (1995). *The Adapted Mind : Evolutionary Psychology and the Generation of Culture*. Oxford University Press, USA.

[Barros, 2007] Barros, B. (2007). Group Size, Heterogeneity, and Prosocial Behavior: Designing Legal Structures to Facilitate Cooperation in a Diverse Society. *SSRN eLibrary*.

[Bazzan et al., 2008] Bazzan, A. L. C., Dahmen, S. R., and Baraviera, A. T. (2008). Simulating the effects of sanctioning for the emergence of cooperation in a public goods game. In *AAMAS '08*, pages 1473–1476, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

[Becker, 1968] Becker, G. S. (1968). Crime and punishment: An economic approach. *The Journal of Political Economy*, 76(2):169–217.

[Bicchieri, 2006] Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, New York.

[Bicchieri and Chavez, 2010] Bicchieri, C. and Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making*, 23(2):161–178.

[Bicchieri and Xiao, 2007] Bicchieri, C. and Xiao, E. (2007). Do the right thing: But only if others do so. MPRA Paper 4609, University Library of Munich, Germany.

[Bickel and Johnson, 2003] Bickel, W. K. and Johnson, M. W. (2003). *Time and decision*, chapter Delay discounting: a fundamental behavioral process of drug dependence. New York: Russell Sage Foundation.

[Blanc et al., 2005] Blanc, A., Liu, Y.-K., and Vahdat, A. (2005). Designing incentives for peer-to-peer routing. In *INFOCOM*, pages 374–385.

[Bo and Bo, 2009] Bo, E. D. and Bo, P. D. (2009). "do the right thing:" the effects of moral suasion on cooperation. NBER Working Papers 15559, National Bureau of Economic Research, Inc.

[Boella et al., 2009] Boella, G., Caire, P., and van der Torre, L. (2009). Norm negotiation in online multi-player games. *Knowl. Inf. Syst.*, 18:137–156.

[Boella and Lesmo, 2002] Boella, G. and Lesmo, L. (2002). A game theoretic approach to norms and agents. *Cognitive Science Quarterly*, pages 492–512.

[Boella et al., 2008] Boella, G., Torre, L., and Verhagen, H. (2008). Introduction to the special issue on normative multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 17(1):1–10.

[Bowling and Veloso, 2002] Bowling, M. H. and Veloso, M. M. (2002). Multiagent learning using a variable learning rate. *Artificial Intelligence*, 136(2):215–250.

[Boyd and Richerson, 1992] Boyd, R. and Richerson, P. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13(3):171–195.

[Brennan et al., 2008] Brennan, G., Gonzlez, L. G., Gth, W., and Levati, M. V. (2008). Attitudes toward private and collective risk in individual and strategic choice situations. *Journal of Economic Behavior & Organization*, 67(1):253 – 262.

[Brito et al., 2009] Brito, I., Pinyol, I., Villatoro, D., and Sabater-Mir, J. (2009). Hiherei: human interaction within hybrid environments regulated through electronic institutions. In *AAMAS '09: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, pages 1417–1418, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

[Campennì et al., 2009] Campennì, M., Andrighetto, G., Cecconi, F., and Conte, R. (2009). Normal = normative? the role of intelligent agents in norm innovation. *Mind & Society*, 8:153–172.

[Cardoso and Oliveira, 2008] Cardoso, H. L. and Oliveira, E. (2008). Electronic institutions for b2b: dynamic normative environments. *Artif. Intell. Law*, 16:107–128.

[Cardoso and Oliveira, 2009] Cardoso, H. L. and Oliveira, E. (2009). Adaptive deterrence sanctions in a normative framework. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 02*, WI-IAT '09, pages 36–43, Washington, DC, USA. IEEE Computer Society.

[Casari, 2004] Casari, M. (2004). On the design of peer punishment experiments. UFAE and IAE Working Papers 615.04, Unitat de Fonaments de l'Anlisi Econmica (UAB) and Institut d'Anlisi Econmica (CSIC).

[Castelfranchi and Conte, 1995] Castelfranchi, C. and Conte, R. (1995). *Artificial societies: The computer simulation of social life*, chapter Simulative understanding of norm functionalities in social groups, pages 252–267. UCL Press.

[Castelfranchi et al., 2000] Castelfranchi, C., Dignum, F., Jonker, C. M., and Treur, J. (2000). Deliberative normative agents: Principles and architecture. In *6th International Workshop on Intelligent Agents VI, Agent Theories, Architectures, and Languages (ATAL),*, pages 364–378, London, UK. Springer-Verlag.

[Chao et al., 2008] Chao, I., Ardaiz, O., and Sanguesa, R. (2008). Tag mechanisms evaluated for coordination in open multi-agent systems. pages 254–269.

[Cialdini and Goldstein, 2004] Cialdini, R. and Goldstein, N. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591– 621.

[Cialdini et al., 1990] Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015–1026.

[Cohen and Levesque, 1990] Cohen, P, R. and Levesque, H., J. (1990). Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213–261.

[Coleman, 1998] Coleman, J. (1998). Foundations of social theory.

[Conte, 2009] Conte, R. (2009). Rational, goal-oriented agents. In *Encyclopedia of Complexity and Systems Science*, pages 7533–7548.

[Conte and Castelfranchi, 1995] Conte, R. and Castelfranchi, C. (1995). *Cognitive and social action*. University College of London Press, London.

[Conte and Castelfranchi, 1999] Conte, R. and Castelfranchi, C. (1999). From conventions to prescriptions. towards a unified theory of norms. *AI and Law*, 7:323–340.

[Conte and Castelfranchi, 2006] Conte, R. and Castelfranchi, C. (2006). The mental path of norms. *Ratio Juris*, 19(4):501–517.

[Conte and Paolucci, 2002] Conte, R. and Paolucci, M. (2002). *Reputation in artificial societies: Social beliefs for social order*. Kluwer Academic Publishers.

[Croson, 2005] Croson, R. (2005). The method of experimental economics. *International Negotiation*, 10:131–148.

[Curbera et al., 2002] Curbera, F., Duftler, M., Khalaf, R., Nagy, W., Mukhi, N., and Weerawarana, S. (2002). Unraveling the web services web: An introduction to soap, wsdl, and uddi. *IEEE Internet Computing*, 6:86–93.

[de Melo et al., 2011] de Melo, C., Carnevale, P., and Gratch, J. (2011). The effect of expression of anger and happiness in computer agents on negotiations with humans. In *Tenth International Conference on Autonomous Agents and Multiagent Systems*.

[de Pinninck et al., 2007a] de Pinninck, A. P., Sierra, C., and Schorlemmer, M. (2007a). Distributed norm enforcement via ostracism. In *COIN @ MALLOW*.

[de Pinninck et al., 2007b] de Pinninck, A. P., Sierra, C., and Schorlemmer, W. M. (2007b). Friends no more: Norm enforcement in multi-agent systems. In Durfee, E. H.; Yokoo, M., editor, *Proceedings of AAMAS 2007*, pages 640–642.

[de Pinninck Bas et al., 2010] de Pinninck Bas, A. P., Sierra, C., and Schorlemmer, M. (2010). A multiagent network for peer norm enforcement. *Autonomous Agents and Multi Agent Systems*, 21:397–424.

[de Vos et al., 2001] de Vos, H., Smaniotto, R., and Elsas, D. (2001). Reciprocal altruism under conditions of partner selection. *Ration Soc*, 13(2):139–183.

[Decker et al., 1997] Decker, K., Sycara, K. P., and Williamson, M. (1997). Middle-agents for the internet. In *IJCAI (1)*, pages 578–583.

[Delgado, 2002] Delgado, J. (2002). Emergence of social conventions in complex networks. *Artificial Intelligence*, 141(1-2):171–185.

[Delgado et al., 2003] Delgado, J., Pujol, J. M., and Sangüesa, R. (2003). Emergence of coordination in scale-free networks. *Web Intelli. and Agent Sys.*, 1(2):131–138.

[Dignum et al., 2000] Dignum, F., Morley, D., Sonenberg, E., and Cavedon, L. (2000). Towards socially sophisticated bdi agents. In *Proceedings of the Fourth International Conference on MultiAgent Systems (ICMAS-2000)*, pages 111–, Washington, DC, USA. IEEE Computer Society.

[Dreber et al., 2008] Dreber, A., Rand, D. G., Fudenberg, D., and Nowak, M. A. (2008). Winners don't punish. *Nature*, 452(7185):348–351.

[Elster, 1989] Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives 3*, 4:99–117.

[Epstein, 2000] Epstein, J. M. (2000). Learning to be thoughtless: Social norms and individual computation. Working Papers 00-03-022, Santa Fe Institute.

[Esteva, 2003] Esteva, M. (2003). *Electronic Institutions: from specification to development*. IIIA PhD Monography. Vol. 19.

[Esteva et al., 2008] Esteva, M., Rodriguez-Aguilar, J. A., Arcos, J.-L., Sierra, C., Noriega, P., Rosell, B., and de la Cruz, D. (2008). Electronic institutions development environment (demo paper). In *Proc. of AAMAS'08*.

[Farrell and Rabin, 1996] Farrell, J. and Rabin, M. (1996). Cheap talk. *The Journal of Economic Perspectives*, 10(3):pp. 103–118.

[Fehr and Gachter, 2000] Fehr, E. and Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

[Fehr and Gachter, 2002] Fehr, E. and Gachter, S. (2002). Altruistic punishment in humans. *Nature*, 415:137–140.

[Fischbacher, 2006] Fischbacher, U. (2006). Abstract z-tree: Zurich toolbox for ready-made economic experiments.

[Fowler and Christakis, 2010] Fowler, J. H. and Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences*, 107(12):5334–5338.

[Freeman, 1979] Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):215–239.

[Fudenberg and Levine, 1998] Fudenberg, D. and Levine, D. K. (1998). *The Theory of Learning in Games*. The MIT Press.

[Galbiati and D'Antoni, 2007] Galbiati, R. and D'Antoni, M. (2007). A signalling theory of nonmonetary sanctions. *International Review of Law and Economics*, 27:204–218.

[Garcia-Camino A, 2006] Garcia-Camino A, Rodriguez-Aguilar A, S. C. V. W. (2006). Norm-oriented programming of electronic institutions. In *AAMAS '06*.

[Giardini et al., 2010] Giardini, F., Andrighetto, G., and Conte, R. (2010). A cognitive model of punishment. In *COGSCI 2010, Annual Meeting of the Cognitive Science Society 11-14 August 2010,*, pages –. Portland, Oregon.

[Gintis, 2003] Gintis, H. (2003). The Hitchhiker's Guide to Altruism: Gene-culture Coevolution, and the Internalization of Norms. *Journal of Theoretical Biology*, 220(4):407–418.

[Gintis, 2004] Gintis, H. (2004). The genetic side of gene-culture coevolution: internalization of norms and prosocial emotions. *Journal of Economic Behavior and Organization*, 53:57–67.

[Goodall, 2005] Goodall, C. E. (2005). *Modifying Smoking Behavior Through Public Service Announcements and Cigarette Package Warning Labels: A Comparison of Canada and the United States.* PhD thesis, Ohio State.

[Griffiths and Luck, 2010] Griffiths, N. and Luck, M. (2010). Changing neighbours: Improving tag-based cooperation. In *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems*, pages 249–256.

[Grizard et al., 2006] Grizard, A., Vercouter, L., Stratulat, T., and Muller, G. (2006). A peer-to-peer normative system to achieve social order. In *Workshop on COIN @ AAMAS' 06*.

[Grossi et al., 2007] Grossi, D., Aldewereld, H. M., and Dignum, F. (2007). Ubi lex, ibi poena: Designing norm enforcement in e-institutions. In *Coordination, Organizations, Institutions, and Norms in Agent Systems II*, pages 101–114. Springer.

[Grossklags, 2007] Grossklags, J. (2007). Experimental economics and experimental computer science: a survey. In *Experimental computer science on Experimental computer science*, pages 12–12, Berkeley, CA, USA. USENIX Association.

[Grusec and Kuczynski, 1997] Grusec, J. E. and Kuczynski, L. (1997). *Parenting and children's internalization of values: a handbook of contemporary theory*. New York: Wiley.

[Hales, 2002] Hales, D. (2002). Group reputation supports beneficent norms. *Journal of Artificial Societies and Social Simulation*, 5.

[Hales and Arteconi, 2006] Hales, D. and Arteconi, S. (2006). Slacer: A self-organizing protocol for coordination in p2p networks. *IEEE Intelligent Systems*, 21:29,35.

[Hales and Edmonds, 2005] Hales, D. and Edmonds, B. (2005). Applying a socially-inspired technique (tags) to improve cooperation in p2p networks. *EEE Transactions in Systems, Man and Cybernetics - Part A: Systems and Humans, 35*, 3:385–395.

[Haley, 2003] Haley, D. F. . K. J. (2003). *Genetic and cultural evolution of cooperation*, chapter The strategy of affect: emotions in human cooperation, pages 7–36. Cambridge, MA: The MIT Press.

[Hardin, 1968] Hardin, G. (1968). The tragedy of the commons. *Science*, xx:1243–47.

[Hardin, 1971] Hardin, R. (1971). Collective action as an agreeable n-prisoners' dilemma. *Behavioral Science*, 16(5):472–481.

[Heckathorn, 1996] Heckathorn, D. D. (1996). The dynamics and dilemmas of collective action. *American Sociological Review*, 61(2):250–277.

[Hirschman, 1984] Hirschman, A. O. (1984). Against parsimony: Three easy ways of complicating some categories of economic discourse. *American Economic Review*, 74(2):89–96.

[Holland, 1993] Holland, J. (1993). The effects of labels (tags) on social interactions. *Working Paper Santa Fe Institute*, 93-10-064.

[Horne, 2003] Horne, C. (2003). The internal enforcement of norms. *Eur Sociol Rev*, 19(4):335–343.

[Houser and Xiao, 2010] Houser, D. and Xiao, E. (2010). Understanding context effects. *Journal of Economic Behavior & Organization*, 73(1):58–61.

[Joseph and Prakken, 2009] Joseph, S. and Prakken, H. (2009). Coherence-driven argumentation to norm consensus. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, ICAIL '09, pages 58–67, New York, NY, USA. ACM.

[Kelsen, 1979] Kelsen, H. (1979). *General Theory of Norms*. Hardcover.

[Kittock, 1993] Kittock, J. E. (1993). Emergent conventions and the structure of multi-agent systems. In *Lectures in Complex systems: the proceedings of the 1993 Complex systems summer school, Santa Fe Institute Studies in the Sciences of Complexity Lecture Volume VI, Santa Fe Institute*, pages 507–521. Addison-Wesley.

[Kittock, 1994] Kittock, J. E. (1994). The impact of locality and authority on emergent conventions: initial observations. In *Proceedings of AAAI'94*, volume 1, pages 420–425. American Association for Artificial Intelligence.

[Kollock and Smith, 1996] Kollock, P. and Smith, M. (1996). Managing the Virtual Commons: Cooperation and Conflict in Computer Communities. pages 109–128. John Benjamins, Amsterdam.

[Lakkaraju and Gasser, 2008] Lakkaraju, K. and Gasser, L. (2008). Norm emergence in complex ambiguous situations. In *Proceedings of the Workshop on Coordination, Organizations, Institutions and Norms at AAAI*.

[Lazer et al., 2009] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., and Van Alstyne, M. (2009). Social science: Computational social science. *Science*, 323(5915):721–723.

[Lotzmann and Möhring, 2009] Lotzmann, U. and Möhring, M. (2009). Simulating norm formation: an operational approach. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '09, pages 1323–1324, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

[Luck et al., 2005] Luck, M., McBurney, P., Shehory, O., and Willmott, S. (2005). *Agent Technology: Computing as Interaction (A Roadmap for Agent Based Computing)*. AgentLink.

[MacNorms, 2008] MacNorms (2008). *MacNorms: Mechanisms for Self Organization and Social Control generators of Social Norms*. http://www.iiia.csic.es/es/project/macnorms-0.

[Masclet, 2003] Masclet, D. (2003). L'analyse de l'influence de la pression des pairs dans les quipes de travail. CIRANO Working Papers 2003s-35, CIRANO.

[Masclet et al., 2003] Masclet, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1):366–380.

[Matlock and Sen, 2009] Matlock, M. and Sen, S. (2009). Effective tag mechanisms for evolving cooperation. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '09, pages 489–496, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

[Mead, 1963] Mead, M. (1963). *Cultural patterns and technical change.* New York: The New American Library.

[Meneguzzi and Luck, 2009] Meneguzzi, F. and Luck, M. (2009). Norm-based behaviour modification in bdi agents. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '09, pages 177–184, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

[Mukherjee et al., 2007] Mukherjee, P., Sen, S., and Airiau, S. (2007). Norm emergence in spatially contrained interactions. In *Proceedings of ALAg-07*, Honolulu, Hawaii, USA.

[Mukherjee et al., 2008] Mukherjee, P., Sen, S., and Airiau, S. (2008). Norm emergence under constrained interactions in diverse societies. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 2*, AAMAS '08, pages 779–786, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

[Mungovan et al., 2011] Mungovan, D., Howley, E., and Duggan, J. (2011). The influence of random interactions and decision heuristics on norm evolution in social networks. *Computational & Mathematical Organization Theory*, pages 1–27. 10.1007/s10588-011-9085-7.

[Newman, 2003] Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM REVIEW*, 45:167–256.

[Nikiforakis and Normann, 2008] Nikiforakis, N. and Normann, H.-T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4):358–369.

[Noriega, 1997] Noriega, P. (1997). *Agent-Mediated Auctions: The Fishmarket Metaphor*. IIIA Phd Monography. Vol. 8.

[Nosratinia et al., 2007] Nosratinia, A., Member, S., and Member, T. E. H. (2007). Grouping and partner selection in cooperative wireless networks. *IEEE Journal on Selected Areas in Communications*, 25:369–378.

[Noussair et al., 2003] Noussair, C., Masclet, D., Tucker, S., and Villeval, M. (2003). Monetary and non-monetary punishment in the voluntary contributions mechanism. Open access publications from tilburg university, Tilburg University.

[Noussair and Tucker, 2005a] Noussair, C. and Tucker, S. (2005a). Combining monetary and social sanctions to promote cooperation. Open Access publications from Tilburg University urn:nbn:nl:ui:12-377935, Tilburg University.

[Noussair and Tucker, 2005b] Noussair, C. and Tucker, S. (2005b). Combining monetary and social sanctions to promote cooperation. *Economic Inquiry*, 43(3):649–660.

[Nowak and Sigmund, 2005] Nowak, M. A. and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature.*, 437(7063):1291–1298.

[Ohtsuki and Iwasa, 2004] Ohtsuki, H. and Iwasa, Y. (2004). How should we define goodness??reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology*, 231(1):107–120.

[Ostrom, 2000] Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14(3):137–158.

[Ostrom et al., 1992] Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *The American Political Science Review*, 86(2):404–417.

[Papaioannou and Stamoulis, 2005] Papaioannou, T. and Stamoulis, G. (2005). An incentives' mechanism promoting truthful feedback in peer-to-peer systems. In *Cluster Computing and the Grid, 2005. CCGrid 2005. IEEE International Symposium on*, volume 1, pages 275 – 283 Vol. 1.

[Parsons, 1937] Parsons, T. (1937). *The structure of social action. A study in social theory with special reference to a group of recent European writers*. New York, London: Free Press.

[Pasquier et al., 2006] Pasquier, P., Flores, R. A., and Chaib-draa, B. (2006). An ontology of social control tools. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, AAMAS '06, pages 1369–1371, New York, NY, USA. ACM.

[Posner and Rasmusen, 1999] Posner, R. and Rasmusen, E. (1999). Creating and enforcing norms, with special reference to sanctions. Law and Economics 9907004, EconWPA.

[Pujol et al., 2005] Pujol, J. M., Delgado, J., Sangüesa, R., and Flache, A. (2005). The role of clustering on the emergence of efficient social conventions. In *Proceedings of the 19th international joint conference on Artificial intelligence*, pages 965–970, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[Rachlin, 2000] Rachlin, H. (2000). *The science of self-control*. Cambridge, London: Harvard University Press.

[Radwanick, 2010] Radwanick, S. (2010). Facebook captures top spot among social networking sites in india. comScore Inc. Press Release.

[Riolo et al., 2001] Riolo, R., Cohen, M., and Axelrod, R. (2001). Cooperation without reciprocity. *Nature*, 414:441–443.

[Rodríguez-Aguilar, 2003] Rodríguez-Aguilar, J. A. (2003). *On the design and construction of agent-mediated electronic institutions*. IIIA Phd Monography. Vol. 14.

[Saam and Harrer, 1999] Saam, N. J. and Harrer, A. (1999). Simulating norms, social inequality, and functional change in artificial societies. *Journal of Artificial Societies and Social Simulation*, 2(1).

[Sabater and Sierra, 2005] Sabater, J. and Sierra, C. (2005). Review on computational trust and reputation models. *Artif. Intell. Rev.*, 24:33–60.

[Sabater-Mir et al., 2007] Sabater-Mir, J., Pinyol, I., Villatoro, D., and Cuni, G. (2007). Towards hybrid experiments on reputation mechanisms: Bdi agents and humans in electronic institutions. In *Proc. of CAEPIA'07*.

[Sacks et al., 2009] Sacks, A., Levi, M., and Tyler, T. (2009). Conceptualizing legitimacy: Measuring legitimating beliefs. *American Behavioral Scientist.*

[Salazar et al., 2010] Salazar, N., Rodriguez-Aguilar, J. A., and Arcos, J. L. (2010). Robust coordination in large convention spaces. *AI Commun.*, 23:357–372.

[Salazar-Ramirez et al., 2008] Salazar-Ramirez, N., Rodríguez-Aguilar, J. A., and Arcos, J. L. (2008). An infection-based mechanism for self-adaptation in multi-agent complex networks. pages 161–170.

[Savarimuthu et al., 2007a] Savarimuthu, B., Purvis, M., Cranefield, S., and Purvis, M. (2007a). How do norms emerge in multi-agent societies? Mechanisms design. *The Information Science Discussion Paper*, (1).

[Savarimuthu et al., 2007b] Savarimuthu, B. T. R., Cranefield, S., Purvis, M., and Purvis, M. (2007b). Norm emergence in agent societies formed by dynamically changing networks. In *Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, IAT '07, pages 464–470, Washington, DC, USA. IEEE Computer Society.

[Scholtes, 2010] Scholtes, I. (2010). Distributed creation and adaptation of random scale-free overlay networks. In *Proceedings of the 2010 Fourth IEEE International Conference on Self-Adaptive and Self-Organizing Systems*, SASO '10, pages 51–63, Washington, DC, USA. IEEE Computer Society.

[Scott, 2000] Scott, J. (2000). *Social Network Analysis: A Handbook*. Sage Publications, second. edition.

[Sen and Airiau, 2007] Sen, S. and Airiau, S. (2007). Emergence of norms through social learning. *Proceedings of IJCAI-07*, pages 1507–1512.

[Shoham and Tennenholtz, 1992] Shoham, Y. and Tennenholtz, M. (1992). On the synthesis of useful social laws for artificial agent societies (preliminary report). In *Proceedings of the AAAI Conference*, pages 276–281.

[Shoham and Tennenholtz, 1994] Shoham, Y. and Tennenholtz, M. (1994). Co-learning and the evolution of social acitivity. Technical report, Stanford, CA, USA.

[Shoham and Tennenholtz, 1997a] Shoham, Y. and Tennenholtz, M. (1997a). On the emergence of social conventions: modeling, analysis, and simulations. *Artificial Intelligence*, 94:139–166.

[Shoham and Tennenholtz, 1997b] Shoham, Y. and Tennenholtz, M. (1997b). On the emergence of social conventions: Modeling, analysis, and simulations. *Journal of Artificial Intelligence*, 94(1-2):139–166.

[Sierra et al., 2004] Sierra, C., Rodríguez-Aguilar, J., Noriega, P., Arcos, J., and Esteva, M. (2004). Engineering multi-agent systems as electronic institutions. *Upgrade*, 5:33–38.

[Sigmund, 2007] Sigmund, K. (2007). Punish or perish? retaliation and collaboration among humans. *Trends Ecol Evol*, 22(11):593–600.

[Sunstein, 1996] Sunstein, C. R. (1996). Social norms and social roles. *Columbia Law Review*, 96(4):903–968.

[Tinnemeier et al., 2010] Tinnemeier, N., Dastani, M., and Meyer, J.-J. (2010). Programming norm change. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1 - Volume 1*, AAMAS '10, pages 957–964, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

[Toivonen et al., 2009] Toivonen, R., Castelló, X., Eguíluz, V. M., Saramäki, J., Kaski, K., and San Miguel, M. (2009). Broad lifetime distributions for ordering dynamics in complex networks. *Physical Review*, 79.

[Tosto et al., 2006] Tosto, G. D., Paolucci, M., and Conte, R. (2006). Altruism among simple and smart vampires. *International Journal of Cooperative Information Systems.*

[Traxler and Winter, 2009] Traxler, C. and Winter, J. (2009). Survey evidence on conditional norm enforcement. Working Paper Series of the Max Planck Institute for Research on Collective Goods $2009_03, MaxPlanckInstituteforResearchonCollectiveGoods$.

[Tyran and Feld, 2006] Tyran, J.-R. and Feld, L. P. (2006). Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics*, 108(1):135–156.

[Ullman-Margalit, 1977] Ullman-Margalit, E. (1977). *The Emergence of Norms*. Clarendon Press, Oxford.

[Urbano et al., 2009a] Urbano, P., Balsa, J., Antunes, L., and Moniz, L. (2009a). Force versus majority: A comparison in convention emergence efficiency. pages 48–63.

[Urbano et al., 2009b] Urbano, P., Balsa, J. a., Ferreira, Jr., P., and Antunes, L. (2009b). How much should agents remember? the role of memory size on convention emergence efficiency. In *Proceedings of the 14th Portuguese Conference on Artificial Intelligence: Progress in Artificial Intelligence*, EPIA '09, pages 508–519, Berlin, Heidelberg. Springer-Verlag.

[Villatoro et al., 2009] Villatoro, D., Sen, S., and Sabater, J. (2009). Topology and memory effect on convention emergence. In *Proceedings of the International Conference of Intelligent Agent Technology (IAT)*. IEEE Press.

[Villatoro et al., 2010] Villatoro, D., Sen, S., and Sabater-Mir, J. (2010). Of social norms and sanctioning: A game theoretical overview. *International Journal of Agent Technologies and Systems*, 2:1–15.

[von Wright, 1963] von Wright, G. H. (1963). *Norm and Action. A Logical Inquiry*. Routledge and Kegan Paul, London.

[Vygotskii and Cole, 1978] Vygotskii, L. S. and Cole, M. (1978). *Mind in society : the development of higher psychological processes / L. S. Vygotsky ; edited by Michael Cole ... [et al.]*. Harvard University Press, Cambridge :.

[Walker and Wooldridge, 1995] Walker, A. and Wooldridge, M. (1995). Understanding the emergence of conventions in multi-agent systems. In Lesser, V., editor, *Proceedings of the First International Conference on Multi–Agent Systems*, pages 384–389, San Francisco, CA. MIT Press.

[Walker A, 1995] Walker A, Wooldridge, M. (1995). Understanding the emergence of conventions in multi-agent systems. In *Proceedings of ICMAS (International Joint Conference on Multi Agent Systems) (San Francisco)*.

[Watkins and Dayan, 1992] Watkins, C. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3-4):279–292.

[Xiao and Hauser, 2009] Xiao, E. and Hauser, D. (2009). Avoiding the sharp tongue: Anticipated written messages promote fair economic exchange. *Journal of Economic Psychology*, 30(3).

[Xiao and Houser, 2005] Xiao, E. and Houser, D. (2005). Emotion expression in human punishment behavior. *Proc Natl Acad Sci U S A*, 102(20):7398–7401.

[Young, 1993] Young, H. P. (1993). The evolution of conventions. *Econometrica*, 61(1):57–84.

[Young, 2007] Young, H. P. (2007). Social norms. Economics Series Working Papers 307, University of Oxford, Department of Economics.

[Young, 2008] Young, H. P. (2008). Social norms. *The New Palgrave Dictionary of Economics*.

[Zhang and Huang, 2006] Zhang, H. and Huang, S. Y. (2006). Dynamic control of intention priorities of human-like agents. In *Proceeding of the 2006 conference on ECAI 2006*, pages 310–314, Amsterdam, The Netherlands, The Netherlands. IOS Press.