Informatica e diritto, XXXIX annata, Vol. XXII, 2013, n. 1, pp. 99-113

How Social Norms Can Make the World More Regular and Better

FEDERICO CECCONI, GIULIA ANDRIGHETTO, ROSARIA CONTE*

SUMMARY: *1. Introduction – 2. Normative Agents – 3. EMIL-A – 3.1. Norm Recognition Module – 4. Three Colored World – 5. Results – 6. Final Remarks and Discussion*

1. INTRODUCTION

Is there any difference between social norms and mere regularities emerging spontaneously from the behaviours of entities that have no norm-based cognition? And if so, which effects do we expect to observe in a world in which agents are endowed with such a type of cognition? In other words, what type of regularities, if any, does such normative cognition establish?

The scientific literature on the subject matter encourages investigation. In a six-year study conducted at the University of Virginia, Turner and collaborators[1] found that exposing college students to information that corrected misperceptions about campus drinking patterns resulted in dramatic reductions in alcohol-related negative consequences[2].

There are different models describing what a social norm is, its properties, modalities of creation and diffusion. They are essentially inspired to two main directions of thought based on two unrelated notions, regularities and obligations. Regularities, or behavioural norms, are spontaneously emerging phenomena[3]. Obligations, or institutional norms, are deliberately

* F. Cecconi is a senior researcher at the Institute of Cognitive Sciences and Technologies, National Research Council of Italy (ISTC-CNR), Rome; G. Andrighetto is a researcher at ISTC-CNR and at the European University Institute (EUI), Florence; R. Conte is a research director at ISTC-CNR and head of the Laboratory of Agent Based Social Simulation (LABSS).

[1] J. TURNER, H. WESLEY PERKINS, J. BAUERLE, *Declining Negative Consequences Related to Alcohol Misuse Among Students Exposed to a Social Norms Marketing Intervention on a College Campus*, in "Journal of American College Health", 2008, n. 57, pp. 85-93.

[2] R.B. CIALDINI, R.R. RENO, C.A. KALLGREN, *A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places*, in "Journal of Personality and Social Psychology", Vol. 58, 1990, n. 6, pp. 1015-1026.

[3] D. LEWIS, *Convention: A Philosophical Study*, Cambridge, Cambridge University Press, 1969; H.P. YOUNG, *The Evolution of Conventions*, in "Econometrica", n. 61, 1993, pp. 57-84.

issued prescriptions[4]. Behavioural norms are often found in the pro-social variant, or in the statistical variant, as frequent, normal behaviours. Institutional norms are obligation-based, and collapse on legal norms, issued by specified institutional authorities. Behavioural regularities and institutional obligations are complementary phenomena. None or poor attempt at their integration has been made so far[5]. However, the gap is neither desirable nor inevitable.

In this paper an integrated approach will be proposed, based on mental representations. In this approach, social and legal norms are treated as recognized, represented and reasoned upon prescriptive commands. The main difference between them lies in their

(a) *origin* - spontaneously emerging in social norms, institutionally deliberate in legal norms;
(b) *transmission* - with laws conveyed in written form (*lex posita*), while social norms are only orally or behaviourally communicated; and
(c) *enforcement*, which is unidirectionally effectuated by applying explicitly defined, predictable and equal sanctions in the case of laws, and mutually executed through uncertain and not always predictable forms of social control in the case of social norms.

Although norm addressees perceive these specificities to some extent, the requisites necessary for representing, reasoning and deciding on norms are common to both laws and social prescriptions. Only a theory that explores the impact of norms on the minds of agents can explain the link between these and other typologies – religious, moral, aesthetical and technical – of norms. Necessarily, a theory of this sort will explore a twofold dynamic of norms, which leads to their surfacing in observable behaviours (*emergence*) on the one hand, and to different levels and kinds of mental processing and representation (*immergence*) on the other[6].

[4] H.L.A. HART, *The Concept of Law*, Oxford, Oxford University Press, 1961; H. KELSEN, *General Theory of Norms*, New York, Oxford University Press, 1979; G.H. VON WRIGHT, *Norms and Action*, London, Routledge and Kegan Paul, 1963.

[5] R. CONTE, *L'obbedienza intelligente*, Bari, Laterza, 1998; R. CONTE, C. CASTELFRANCHI, *From Conventions to Prescriptions. Towards an Integrated View of Norms*, in "Artificial Intelligence and Law", Vol. 7, 1999, pp. 119-125; R. CONTE, G. ANDRIGHETTO, M. CAMPENNÌ (eds.), *Minding Norms. Mechanisms and Dynamics of Social Order in Agent Societies*, Oxford Series on Cognitive Models and Architectures, New York, Oxford University Press, Forthcoming.

[6] See G. ANDRIGHETTO, M. CAMPENNÌ, F. CECCONI, R. CONTE, *The Complex Loop of Norm Emergence: A Simulation Model*, in Takadama K., Cioffi-Revilla C., Deffuant G.

This work is based on a computational methodology, i.e., multi-agent-based simulation[7]. This is an ideal tool for exploring the two-way dynamics of norm emergence, because it must explicitly and completely describe the whole process leading from a no-norm world to one in which regularities of some sort exist. In addition, by using agent-based simulation the relationship between cognition and social dynamics can start to be teased apart in a dynamic manner, and their respective roles accounted for.

In this paper, we present agent-based simulations aimed to understand what would happen in a world populated by normative agents, able to recognize norms and to reason upon them, compared to other, cognitively, less complex agents, following only their own individual goals.

2. Normative Agents

The development of normative architectures is a burgeoning research field[8]. However, architectures of normative agents are predominantly inspired in some way by BDI - *Belief-Desire-Intention* architectures, introduced

(eds.), "Simulating Interacting Agents and Social Phenomena", New York, Springer, 2010, pp. 19-35.

[7] N. Gilbert, K.G. Troitzsch, *Simulation for the Social Scientist*, Maidenhead, Open University Press, 2nd ed., 2005; S. Moss, P. Davidsson (eds.), *Multi-Agent-Based Simulation*, LNAI 1979, Berlin, Springer, 2000, pp. 157-166; R. Conte, N. Gilbert, *Computer Simulation for Social Theory*, in Gilbert N., Conte R. (eds.), "Artificial Societies: The Computer Simulation of Social Life", London, UCL Press, 2006, pp. 1-18; J.M. Epstein, *Generative Social Science: Studies in Agent-based Computational Modelling*, Princeton, Princeton University Press, 2006; see also the proceedings of the Multi-Agent-Based Simulation (MABS) Workshops, http://www.pcs.usp.br/ ~ mabs/mabs_intro.html.

[8] R. Conte, G. Andrighetto, M. Campennì (eds.), *op. cit.*; G. Andrighetto, M. Campennì, F. Cecconi, R. Conte, *op. cit.*; J. Broersen, M. Dastani, J. Hulstijn, Z. Huang, L. van der Torre, *The BOID Architecture: Conflicts Between Beliefs, Obligations, Intentions and Desires*, in "Proceedings of the 5th International Conference on Autonomous Agents and Multi Agent Systems (AAMAS)", New York, ACM, 2001, pp. 9-16; N. Criado, E. Argente, V. Botti, P. Noriega, *Reasoning about Norm Compliance*, in "Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems", 2011, pp. 1191-1192; B.T.R. Savarimuthu, S. Cranefield, M.A. Purvis, M.K. Purvis, *Obligation Norm Identification in Agent Societies*, in "Journal of Artificial Societies and Social Simulation", Vol. 13, 2010, n. 4, http://jass.soc.surrey.ac.uk/13/4/3.html; F. Lopez y Lopez, M. Luck, *Modelling Norms for Autonomous Agents*, in Chavez E., Favela J., Mejia M., Oliart A. (eds.), "Fourth Mexican International Conference on Computer Science", IEEE Computer Society, 2003, pp. 238-245.

by the pivotal work of Rao and Georgeff[9], which can be regarded as the point of departure for further developments. The BDI framework is intended to model human intelligent action and decision-making. In the last decade, BDI architectures augmented with obligations, like BOID - *Beliefs-Obligations-Intentions-Desires*[10] or BDOING - *Beliefs, Desires, Obligations, Intentions, Norms and Goals*[11], began to appear.

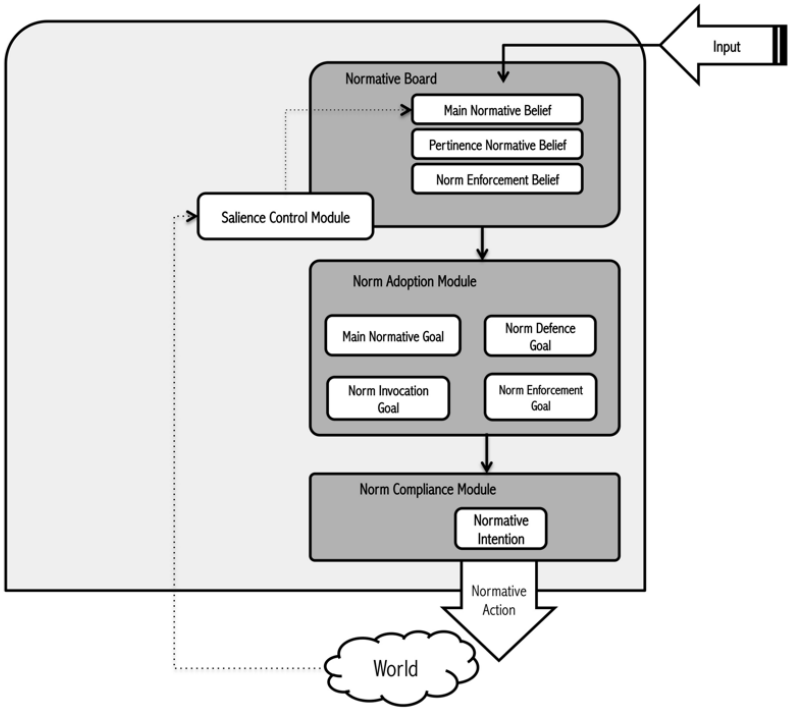


Fig. 1 – Main components and mental dynamics of EMIL-A: the architecture consists of different modules interacting with one another by means of input-output mechanisms

[9] A.S. RAO, M.P. GEORGEFF, *Decision. Procedures for BDI Logics*, in "Journal of Logic and Computation", 1998, n. 8, pp. 293-343.

[10] J. BROERSEN, M. DASTANI, J. HULSTIJN, Z. HUANG, L. VAN DER TORRE, *op. cit.*

[11] F. DIGNUM, D. KINNY, L. SONENBERG, *From Desires, Obligations and Norms to Goals*, in "Cognitive Science Quarterly", Vol. 2, 2002, n. 3-4, pp. 407-430.

The normative architecture we present here, EMIL-A[12], is inspired to BOID and BDOING as it entails the representation of normative beliefs and goals based on obligations.

However, unlike BOID and BDOING, EMIL-A includes a module for norm-recognition allowing agents to process incoming inputs and possibly converting them into norms. This mechanism proves essential when modelling and operationalizing the process of norm immergence[13]. In the next Section a description of the main components and processes of EMIL-A is provided[14].

3. EMIL-A

A sketch of the main components and mental dynamics of EMIL-A is provided in Fig. 1. In particular, it includes:

1. Two types of representations:
   - Normative Beliefs: beliefs that a given behaviour, in a given scenario, for a given set of agents, is either forbidden, obligatory, or permitted[15].
   - Normative Goals: goals[16] relativized to a normative belief. A goal is relativized when it is held because and to the extent that a given world-state or event is hold to be true or is expected[17].
2. Three modules:
   - Norm Recognition, which takes an observed behaviour or a message as an input and possibly turn it into a new normative belief.

[12] This normative architecture has been developed within the EMIL project (EMergence In the Loop: simulating the two way dynamics of norm innovation), a FET-funded European project on the agent-based simulation of the two-way dynamics of norm innovation.

[13] G. ANDRIGHETTO, M. CAMPENNÌ, F. CECCONI, R. CONTE, *op. cit.*

[14] R. CONTE, G. ANDRIGHETTO, M. CAMPENNÌ (eds.), *op. cit.*

[15] R. CONTE, C. CASTELFRANCHI, *The Mental Path of Norms*, in "Ratio Juris", Vol. 19, 2006, n. 4, pp. 501-517.

[16] A goal is here meant in the very general sense derived from cybernetics, i.e., a wanted state of the world triggering and driving actions (see, G.A. MILLER, E. GALANTER, K.H. PRIBRAM, *Plans and the Structure of Behavior*, New York, Henry Holt, 1960; R. CONTE, C. CASTELFRANCHI, *Cognitive and Social Action*, London, UCL Press, 1995; R. CONTE, *Rational, Goal Governed Agents*, in Meyers R.A. (ed.), "Springer Encyclopedia of Complexity and System Science", Berlin, Springer, 2009).

[17] P.R. COHEN, H.J. LEVESQUE, *Persistence, Intention, and Commitment*, in Cohen P.R., Morgan J., Pollack M.A. (eds.), "Intentions in Communication", Cambridge, MIT Press, 1990, pp. 33-71.

- Norm Adoption, which takes a normative belief as an input and possibly gives a new normative goal as an output.
- Norm Compliance; which takes a normative goal as an input and possibly puts it into execution, performing a normative action.

3. The norms' salience mechanism, which updates the salience of norms, according to external events.

The Norm Recognition Module is the crucial component by means of which agents are able to infer that a certain norm is in force even when it is not already stored in their normative memory. Implementing such a capacity is conditioned to modelling agents' ability to recognize an observed or communicated social input as normative. It allows agents to form new normative beliefs processing the information received while interacting with or observing the other agents behaving in a common environment. The Norm Recognition Module detects whether or not the received social input refers to a normative belief already stored in the normative board. In the former case, it will update the salience of the corresponding norm accordingly. In the latter case, it will either form a new normative belief, or simply discard the input.

When a new normative belief is formed, the Norm Recognition Module will send information to the Norm Adoption module. This will use such information to decide whether or not to form the corresponding Normative Goal, based on the norm-adoption rule[18]. The general mechanism by which an autonomous agent adopts external requests, called adoption mechanism, has been described at some length in Conte and Castelfranchi[19]. Here, suffice it to say that an agent (the adopter) will adopt another agent's (i.e., the adoptee's) goal as hers, on condition that she, the adopter, comes to believe that the achievement of the adoptee's goal will increase the chances that she will in turn achieve one of her previous goals.

Finally, the new Normative Goal will be inputed to the Norm Compliance Module. This consists in a decision-making procedure that possibly turns the new goal into an intended Normative Action. The procedure will put the goal to execution unless it is already realised or incompatible with more important goals. In the last two cases, the Normative Goal will be suspended until the conditions for its execution will be verified again.

[18] R. CONTE, C. CASTELFRANCHI, *op. cit.*
[19] *Ibidem*.

In this paper, we will describe the implementation only of the first component of EMIL-A, i.e., the Norm Recognition Module (see Fig. 2). This is most frequently involved in answering the question how a new norm is found out, a topic that we consider particularly crucial in norm emergence, innovation and stabilization.

3.1. Norm Recognition Module

The Norm Recognition Module (see Fig. 2) consists of a normative board (on the left), that is a long-term memory, and a two-layer working memory (on the right). The normative board contains normative beliefs, ordered by salience. By norm salience, we refer to the measure that indicates how active a norm is within a group and in a given context[20]. The working memory is where social inputs are elaborated. Agents observe social inputs and receive messages from one another.

Each input is presented as an ordered vector, consisting of four elements:

1. the source (X), i.e., the agent observed or the agent who sends the message;
2. the action transmitted (α), i.e., the potential norm;
3. the type of input (T): it can consist either in a behaviour (B), i.e., an action or reaction of an agent, or in a communicated message, transmitted through the following holders:
   - assertions (A), i.e., generic sentences pointing to or describing a state of the world;
   - requests (R), i.e., requests of action;
   - deontics (D), partitioning situations between good/acceptable and bad/unacceptable. Deontics are holders for the three modal verbs analysed by von Wright[21] "may", indicating permission, "must", indicating obligation, and "must not", indicating prohibition;

[20] See D. VILLATORO, G. ANDRIGHETTO, R. CONTE, J. SABATER-MIR, *Dynamic Sanctioning for Robust and Cost-efficient Norm Compliance*, in "Proceedings of the 22nd International Joint Conference on Artificial Intelligence", 2011, pp. 414-419; G. ANDRIGHETTO, D. VILLATORO, *Beyond the Carrot and Stick Approach to Enforcement: An Agent-based Model*, in Kokinov B., Karmiloff-Smith A., Nersessian N.J. (eds.), "European Perspectives on Cognitive Science", Sofia, New Bulgarian University Press, 2011 for a detailed description of the norm salience dynamics.

[21] G.H. VON WRIGHT, *op. cit.*

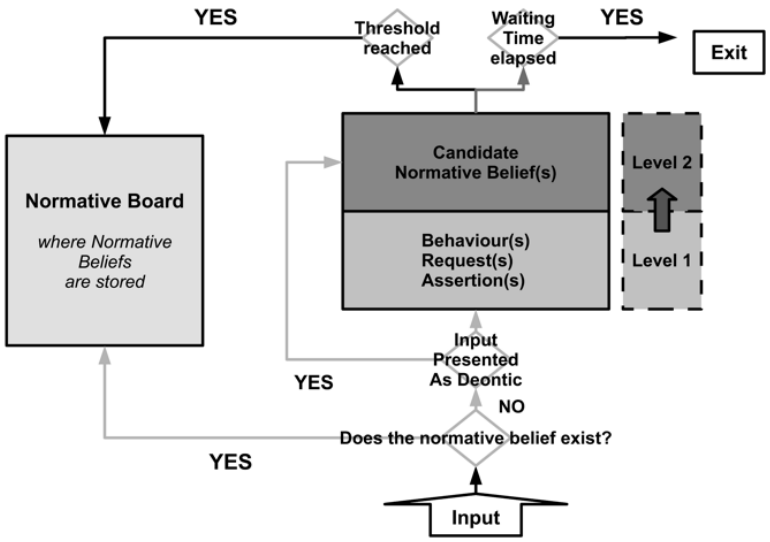


Fig. 2 – The Norm Recognition module

– normative valuations (V), i.e., assertions about what it is right or wrong, correct or incorrect, appropriate or inappropriate (e.g., it is correct to respect the queue).

4. The observer (Y), i.e., the agent who receives the input.

Once the input is received, EMIL-A will compute the information thanks to its norm recognition module. Here follows a brief description of how this module works.

Every time a message containing a deontic (D), for example, "You must answer when asked", or a normative valuation (V), for example "It is impolite not to answer when asked", is received, it will directly access the second layer of the architecture, giving rise to a candidate normative belief "One must answer when asked", which will be temporarily stored. This will sharpen agents' attention: further messages with the same content, especially when observed as open behaviours or transmitted by assertions (A) – for example "When asked, Paul answers" – or requests (R) – for example "Could you answer when asked?" – will be processed and stored at the first level of the architecture. Beyond a certain normative threshold (which represents the frequency of the corresponding normative behaviours observed,

i.e., the percentage of the compliant population), the candidate normative belief will be transformed into a new (real) normative belief, which will be stored in the normative board. The normative threshold can be reached in several ways. For example, by observing a given number of agents performing the same action (α) prescribed by the candidate normative belief, e.g., answering when asked. If the observer receives no further occurrences of the same input (action α), the candidate normative belief will leave the working memory (Exit) after a fixed time t.

Exposed to the normative behaviours of others and to their explicit or implicit normative requests, agents acquire normative beliefs. Normative messages or normative requests alone are not sufficient to generate normative beliefs, they have to be confirmed by the compliant conduct of others, which reveals the actual salience and degree of activity of the norm. Thus for a normative belief to be generated, normative prescriptions have to be transmitted and the correspondent normative actions observed.

In the simulation experiments presented in this work, we have implemented a simplified version of EMIL-A, in which the decision-making is driven only by the indications provided by the Norm Recognition Module.

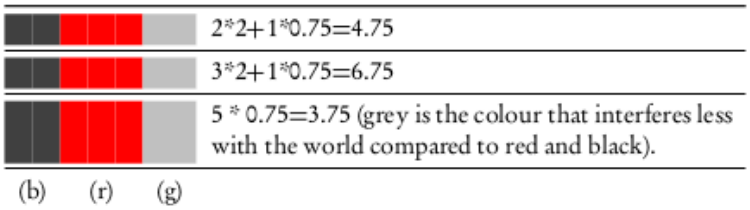
4. Three Colored World

We designed a bidimensional world, divided into regular cells. Agents move on these cells and they can take decisions and modify the states of the cell in which they are (see Fig. 3 and Fig. 4). We refer to this scenario as the three colored world.

In the three colored world there are three kinds of agents (see Fig. 4): (a) agents that are not able to recognize norms, (b) agents able to recognize norms, but at the present stage with no norms able to influence their own behavior (non-active norms), (c) agents endowed with a norm-recognizing mechanism and with active norms.

In this world, agents can *color* the cell of the world where they are, with one of the three colors at their disposal, red, black and grey, and *modify* the propension to follow one of their goals. In particular, the goals agents are endowed with are the following: <goal 1>: minimise interference; <goal 2>: imitate other agents; <goal 3>: normative goal. The propension to follow each of these goals is indicated in terms of probability to use one of the three colors. In details:

– *Goal 1: minimise interference* with the current state of the world. Let us imagine that, starting from a no-color world, each track left on one's cell can produce interference. Within a certain observation window, each agent can count how many cells are already colored by either red, black or grey. Then, depending on the color the cell is left with (which overlays that which was possibly already present in the agent's cell) the interference is calculated as follows: (a) *If RED, interference will be NBlackCells * 2+NGreyCells *0.75, (b) If BLACK, interference will be NRedCells * 2+NGreyCells*0.75, (c) If GREY, interference will be (NredCells+NblackCells)*0.75*. Table 1 shows the three calculations for the same example.



Tab. 1 - Interference of the three colours, black (b), red (r) and grey (g) with the state of the world

– *Goal 2: imitate neighbours*: agents aim to use the same color of their neighbors. In the border between two different color area, this can conflict with the preceding goal.
– *Goal 3*: (for norm detectives only) *use the most salient norm*, which will specify what color to use.

Agents' intelligent decisions are goal based[22]. The type of goals that can be satisfied is the same for all of the agents, but each agent can, at different moments, obey a different goal.

For example, let us imagine four possible goals: go to work, have some sleep, play the piano, go jogging (G1, G2, G3, G4). At each moment, agents

[22] Intelligent agents differ from utilitarian agents as they try to satisfy their most important goals. Goal satisfaction does not imply utility maximization because the advantage of the benefits obtained over the costs sustained to achieve the best goal may be lower than that realized by satisficing the next-best goal option. See R. CONTE, R. PEDONE, *Finding the Best Partner: The PART-NET System*, in "Proceedings of the 1st International Workshop on Multi-Agent Systems and Agent-Based Simulation", 1998, pp. 156-168.

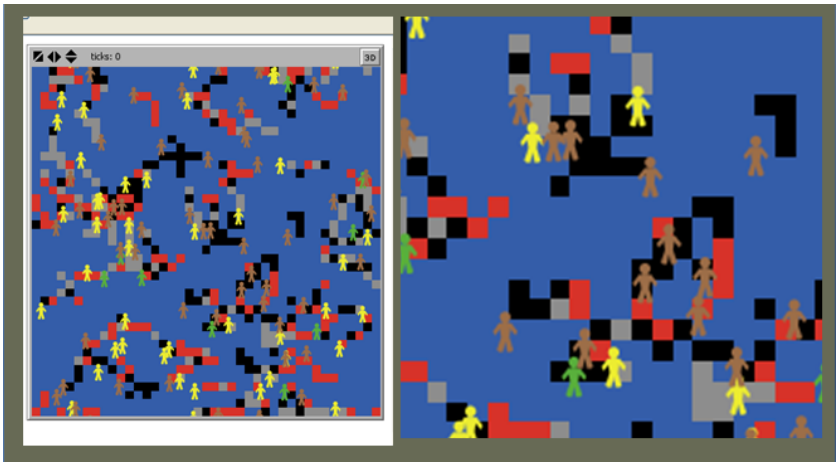


Fig. 3 – Agents and the environment

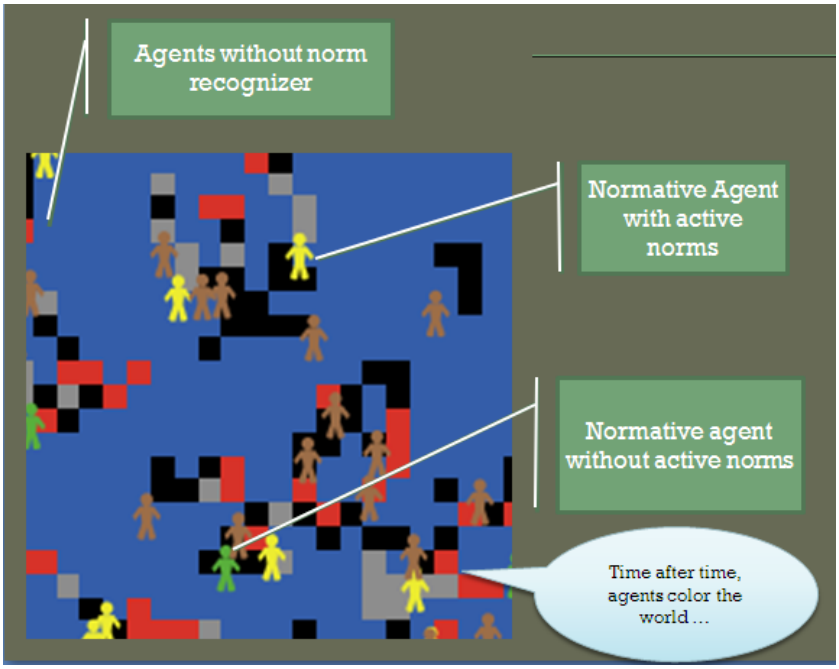


Fig. 4 – Agents' types

Agent	List of probabilities				Behavior
	G1 Work	G2 sleep	G3 play	G4 jogging	
A1	0	0	0	1	With a 100% probability, A1 wants to satisfy the goal to go out jogging. A1 will, thus, follow action b1.
A2	0.5	0	0	0.5	With a 50% probability, A2 wants to satisfy the goal to go to work, and with 50% probability the goal to go out jogging. The selected action is again b1.
A3	0.25	0.25	0.25	0.25	Flip coin.
A4	1	0	0	0	Dual of A1, but the goal to be satisfied is G1, i.e, go to work.
A5	0.9	0.1	0	0	With a 90% probability, A5 wants to satisfy the goal to go to work, and with a 10% probability the goal to have some sleep. A5 is more likely to chose action b1 than anything else.

Tab. 2 - Agents, actions and goals

can choose one out of three possible actions: b1 – going out (useful for satisfying G1 and G4); b2, getting a piano score (useful for satisfying G3); b3, staying home, useful for satisfying G2. The world is populated by five agents, *A1, A2, A3, A4, A5*, each assigned with a list in which the probability of following one of the four goals is indicated:

Agents' actions modify the world and can also modify their own goals. Normative agents can also communicate messages of the form described in the previous Section. Some of those messages contain deontics prescribing the colour to be used "You must use red/black/grey when colouring the world". These messages favour the generation of normative representations. The interaction of the normative representations with the goals results in the

agents' behaviour. The simulation has been implemented using a Netlogo platform.

5. Results

The simulation allows us to observe how the world is colored, depending on the number of normative agents introduced in the system since the beginning. The results are the average of thirty repetitions for each condition.

Fig. 5 shows the percentage of grey cells, i.e, the cells that interfere less with the world, for different percentages of normative agents present in the simulation since the beginning.

Fig. 6 shows an interesting regularity: in a world populated by normative agents, the steady state at the end of simulation is not monocromatic. Using normative agents, the number of steady state with two colors increase as a linear function of the percentage of normative agents. This result in not trivial: in fact, norms "in the sense of coordination rules" cannot explain this kind of regularity. In other word, the norms increase "variety" of the world. Be careful: all the steady state with two colors contain grey. Grey is one of the component of bicolor ending of simulations.

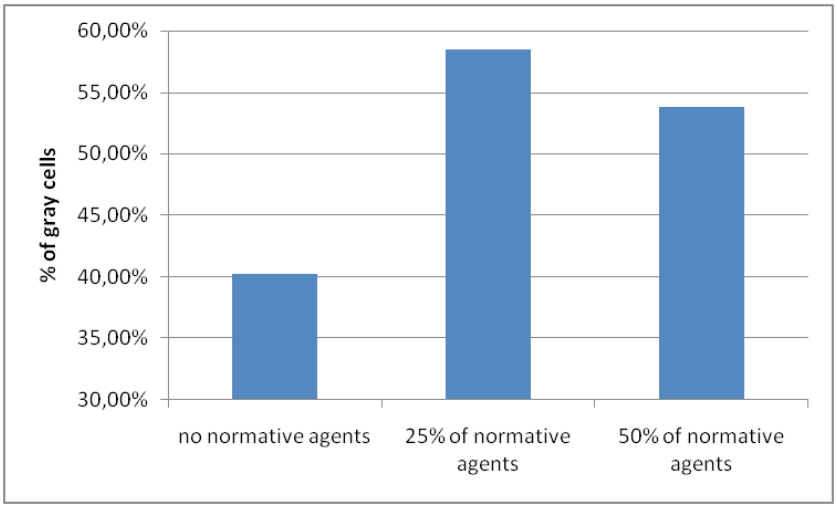


Fig. 5 – Percentage of grey cells at the end of simulations

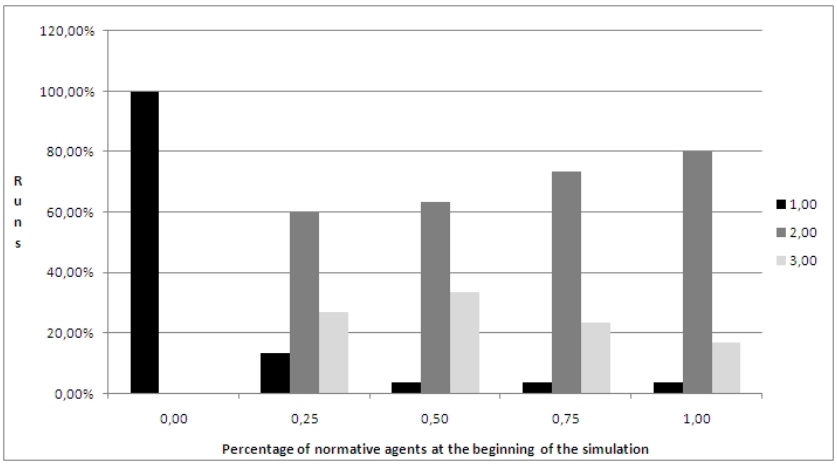


Fig. 6 – The distribution of the colors at the steady state.
Numbers in the legend (1,2,3) indicate the number of colors present in the world.
We show that with no normative agent the world is, at steady state, monocromatic.
On the contrary, using normative agents, the world has two (or three) colors

6. Final Remarks and Discussion

The main result of the experiment concerns the number of grey cells present at the end of the simulation. This varies as a function of the increased number of normative agents: with a non-nul number of agents that are able to recognize norms, the number of grey cells increases. What is more, if normative agents are not around, the simulation converges uniformly on one color, alternatively red, black or grey (the probabilities of occurrence of the three cases directly derive from the degree of grey's interference compared to the other two colors, in our case 0.75 vs 2).

A statistical analysis of the model allows us to conclude that there are no external limits to the convergence on a single color. On the contrary, the presence of normative agents preserves some diversity in the final states.

A second remark is that if the choice of grey is to be preferred to the others for a reason that is external to the model (for example, using the model in a context of social integration we can think that grey is a behavior reducing the probability of contrasts, whilst in an environmental model it may be interpreted as a less disruptive, or more sustainable, behavior), the re-

sults show that the presence of normative agents favors such solution, with interesting hints in the study of policies and, partially, of interventions.

Imitation based on a utility function is enough to bring about convergence and regularity, but does not ensure that such regularity corresponds to a socially desirable result. Regularity is useful when trying to achieve a problem of coordination, as in the case of left or wright precedence. Instead, it is not sufficient with social problems in which solutions are not equivalent, and in which the imitation of individually successive strategies may contrast with the socially preferable solution.