The Immergence of Norms in Agent Worlds

Rosaria Conte, Giulia Andrighetto, Marco Campennì

LABSS - Istituto di Scienze e Tecnologie della Cognizione - CNR, via S. Martino della Battaglia
44, 00185 Rome, Italy e-mail: (http://labss.istc.cnr.it)
{rosaria.conte, giulia.andrighetto, marco.campenni}@istc.cnr.it

Abstract. In this paper, after a short review of the dichotomous view of norms usually seen as either regular behaviors or obligations issued by authorities, norms are proposed to be defined as recognized, represented and reasoned upon prescriptive commands. A normative agent architecture – EMIL-A – is presented and shown to account for a complex bidirectional dynamics of norms as social phenomena that emerge because and to the extent that they immerge in the agents' minds. Simulations run using EMIL-A will be discussed to illustrate the advantages of the present treatment of norms, over either side of the dichotomy.

Introduction

How build up social order in agent worlds? Are we satisfied with current approaches to the emergence of norms? Unfortunately not, as a number of questions are still open. In particular, it is possible to envisage a dichotomy in the scientific treatment of norms. On one hand, social scientists view norms as *regular behaviours*, possibly enforced by social expectations and sanctions. On the other hand, philosophers of law and logicians conceptualise norms as *obligations issued* by definite authorities. Hence the first set of questions, *how keep normative behaviour distinct from normal conduct on one hand and acquiescence under menace on the other*?

As a flexible, adaptive form of social order is crucial in agent worlds, a second set of questions needs to be addressed, i.e. *how do norms emerge, change and get adapted to new circumstances*? In the approach presented here, norms will be conceptualised as social and cognitive phenomena undergoing a complex dynamics, in which the social process of emergence and the mental process of immergence are intertwined in a circular fashion. The interplay between the mental and the social dynamics allows norms to emerge and change. Observable conformity is only the tip of the normative iceberg. The crucial dynamics lies in the minds of the agents, beneath the line of observation. Norms cannot emerge in society unless they previously immerge in the mind, i.e. get converted into mental representations. Agents abiding with norms, or violating them, act on a set of specific, norm-related, mental representations.

The mental dynamics of norms brings about a third set of questions: *how should we characterize the agents from among which norms can emerge*? Current BDI models

of normative agents tackle the questions as to how people represent, reason upon, abide with or violate norms, but they do not address an earlier problem, i.e. how can norms emerge among BDI type of agents.
In this paper, we intend to present an integrated and highly dynamic view of norms showing why and how it enables to answer the three sets of questions listed above, by means of agent based simulation.

1 The Normative Gap

Theories of norms are grounded on two, unrelated notions, *regularities* and *obligations*. Regularities, or behavioural norms, are spontaneously emerging social phenomena. Obligations, or institutional norms, are deliberately issued prescriptions. Behavioural norms are often found in the moral variant, as good or prosocial conduct, 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 of integration has been made so far. However, the gap is neither desirable nor inevitable. In the rest of the paper an integrated approach will be proposed, based on mental representations: social norms, just like legal norms, are recognized, represented and reasoned upon as prescriptive commands. Only a theory that explores the impact of norms on the minds of agents can explain the link between different typologies of norms.
The work presented is based on artificial societies. Agent-based Simulation is an ideal tool for exploring the two-way dynamics of norm emergence because it is relatively free of epistemological assumptions. Thus processes of immergence and emergence can be explored in a way that is very difficult in any other way. In this way the relationship between cognition and social dynamics can start to be teased apart in a truly dynamic manner.

2 A 2-way dynamics of norms

In a view of norms as two-sided, external (social) and internal (mental) objects ([1][2][3][4]), norms come into existence only when they emerge, not only *through* the minds of the agents involved, but also *into* their minds. In other words, they work as norms only when the agents recognize them, reason and take decisions upon them as norms. The emergence of norms implies their immergence into the agents' minds. Only when the normative, i.e. prescriptive, character of a command or other action is recognized by the agent, a norm gives rise to a normative behaviour of that agent. Thus, for a norm-based behavior to take place, a normative belief has to be generated into the minds of the norm addressees, and the corresponding normative goal has to

be formed and pursued. Our claim is that a norm emerges as a norm only when it immerges into the minds of the agents involved; in other words, when agents recognize it as such.

In previous works ([5][6][7][8][9]), we described the process of norm emergence as a gradual and complex dynamics by which the macro-social effect, in our case a specific norm, emerges in the society *while* immerging in the minds of the agents producing it, generating a number of intermediate loops. Thus, before any global effect emerges, specific local events affect the generating systems, their beliefs and goals, in such a way that agents influence one another into converging on one global macroscopic effect. Emergence of social norms is due to the agents' behaviors, but the agents' behaviors are due the the mental mechanisms controlling and (re)producing them (immergence). Of course, our view of norms calls for a cognitive architecture of normative agents, which is not new to the field of agents and multiagent systems (think of the BOID architecture, for example). In the next section, an analysis of our normative architecture, EMIL-A, is presented.

3 Normative agents

In order to model and operationalize the process of norm immergence, autonomous intelligent agents need to be endowed with internal mechanisms and mental representations allowing norms to affect their behaviours. Such representations are commonly realized by architectures inspired to the modular architecture of Artificial Intelligence approaches. Nowadays, there is no unequivocal concept for the design of normative agents. The development of normative architectures is a burgeoning research field. However, architectures of normative agents are predominantly inspired in some way by BDI (Belief-Desire-Intention) architectures, introduced by the pivotal work of Rao and Georgeff ([10]), 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. A particular striking example of this approach is a straightforward extension of the BDI architecture to normative reasoning, denoted as BOID (Belief-Obligations-Intentions-Desires) agent architecture (see [11]), which includes obligations among its mental components.

The normative architecture we present in this section, EMIL-A, is inspired to BOID as it entails the representation of normative beliefs and goals based on obligations. However, unlike BOID, EMIL-A includes a module for norm-recognition, allowing the agent to process incoming inputs and possibly converting them into norms.

3.1 EMIL-A Architecture

Our normative architecture EMIL-A (see [6], for a detailed description) is meant to show that norms not only regulate behaviour but also act on different aspects of the

mind. EMIL-A consists of mechanisms and mental representations allowing agents i) to form normative beliefs and goals, and decide whether to realize them or not and ii) to be more or less reactive to external inputs by means of short cuts. EMIL-A is accessed through the norm recognition module: before an input is recognized as normative, the norm cannot immerge in the minds of agents and, as a consequence, cannot affect their behaviours and emerge in society. We consider existent normative architectures not sufficiently flexible and adaptable to be really plausible; we believe that the future of normative architectures is closely related to the development of hybrid architectures.

3.1.1 Normative mental representations. In this section, we shall endeavour to clarify some components of the mental processing of norms.

Normative belief[1]. First of all, a norm becomes a belief, namely the belief that a given behaviour in a given context for a given set of agents is forbidden, obligatory, permitted, etc. More precisely, the belief should be that "there is a norm prohibiting, prescribing, permitting that..." ([12][13][2][3]). Indeed, norms are aimed at and issued for generating the corresponding beliefs. In other words, norms must be acknowledged as such in order to properly work. Of course, a normative belief does not imply that a given norm has in fact been deliberately issued by some sovereign. Social norms are often set up by virtue of unwanted effects. However, once emerged, a given social norm is believed to be based upon some normative authority, if only an anonymous and impersonal one ("You are wanted, expected (not) to do this…": "It is generally expected that…"; "This is how things are done…", etc.).

Normative Goal. However, a believer is not yet a decider: beliefs are necessary but insufficient conditions for norms to be complied with. What is it that leads agents to accept a norm, which by definition prescribes a costly behaviour?

In the BDI approach intentions and actions originate only from desires. On the contrary, a great deal of our actions are not elicited by our desires but by external pressures and requests. Duties and norms are one of the external sources of our goals. How is this possible? How can norms generate goals?

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 ([14]). Under the effect of social inputs, goals can be generated anew via cognitive factors, as goals *relativized* to other mental states (e.g., social beliefs). 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 ([15]). When goals are positive or pro-social, the process of generation is

1 In EMIL-A, normative beliefs, together with normative goals, are organized and arranged in the normative board according to their respective salience. By *salience* we refer to the norm's degree of activation, which is a function of the number of times a given norm enters the agent's decision-making.

called *goal-adoption* (see [1]). By this means, an autonomous agent (adopter) will have another agent's (adoptee) goal as hers, on condition that she, the adopter, comes to believe that the adoptee's achievement of his goal will increase the chances that the adopter will in turn achieve one of her previous goals. I will probably lend my car to my room mate tonight, if I want to invite my fiancé for dinner.

There seems to be a correspondence between the process from a belief about an ordinary request to the decision of accepting such a request, i.e. the aforementioned process of *social goal adoption*, and the process leading from a normative belief to a normative goal (*norm adoption*): a normative goal of a given agent *x* about action *a* is a goal that *x* happens to have as long as she has a normative belief about *a*. More specifically, *x* has a normative goal only if she believes to be subject to a norm.

Norm Recognizer. The norm recognition module is the main entrance, so to speak, to the EMIL-A architecture. Before an input is recognized as normative, the norm cannot immerge in the minds of agents and, as a consequence, cannot emerge in society. Agents need to be able to discriminate between norms and other social phenomena, such as coercion, ordinary requests, conventions, etc. Our claim is that other normative architectures did not render justice to the recognition procedure (see [9][8]). Simplifying, a given norm is recognized if current input

- matches with a norm already stored in our (normative) memory;
- leads to a new norm being inferred or induced by the agent on the grounds of given indicators.

In the first case, the agent is facilitated by schemata, scripts, or other pragmatic structures ([16][17][18][19]; see [20] for an overview) the norm is embedded in (see [11], for a description). Once these are activated for any reason, the corresponding normative beliefs, expectations and behavioural rules are prompted.

The second option is followed when such scripts, and consequently the corresponding pattern matching operations, are not possible. The agent has no corresponding norm. This is why the norm recognition module is needed. Indeed, the norm-recognizer that we are going to describe tries to answer the question as to how agents tell new norms, not yet stored in their memory (see also [21]). Telling norms implies agents' ability to take an observed or communicated social input as normative, and consequently to form a new normative belief.

EMIL-A module for norm recognition consists of a normative frame by which the received inputs are elaborated and interpreted, and a long term memory - called normative board - where normative beliefs and normative goals once formed are stored and ordered by salience.

The Normative Board. When EMIL-A has to deal with an external input, such as a NO SMOKING sign, the norm recognition module will explore the N-Board. Suppose a corresponding normative belief is found (DO NOT SMOKE WHEN PROHIBITED), a normative belief is fired that will follow the path described previously.

The normative board is an archive in the long term memory where active norms are stored, arranged according to the *salience* gained. Difference in salience has the effect that a subset of norm-related representations interferes more frequently and strongly with the general cognitive processes of the agent. To decide which action to execute, the agent will search through the normative board: if more than one item is found out, the most salient norm will be chosen.

If a norm is never adopted by the agent, its salience begins to decrease, and sooner or later the normative belief will decay. On the contrary, a norm that is frequently processed by the decision-maker, will increase in salience. 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. In such a case, the norm will exit the normative board.

3.2 Value added of EMIL-A

So far, the study of norm emergence has been identified with the study of behavioural regularities. However, not all the regularities are mandatory, and not all the norms are observed. Hence, the logical and pragmatic priority is how agents find out what are the *normative* regularities. Only afterwards, it makes sense to model the reasons why they conform to them. The value added of EMIL-A is to account for this specific aspect of norm-based regulation, how agents find out the norms they decide whether or not to conform to.

Norm recognition is an important requirement of norm-emergence. In previous works (see also [5][7]), emergence has been defined as a gradual and complex dynamics by which the macro-social effect, in our case a specific norm, is brought about in society *while* immerging in the minds of the agents, generating it through a number of intermediate loops.

Unlike moral dispositions, norm-recognition is poorly sensible to subjective variability, and rather robust. It allows us to (a) account for the universal appearance of norms in human and primate societies; (b) render justice to the intuition that humans violate norms, but have little problems in finding them out; (c) account for the evolutionary psychological evidence (see [22][23]) that agents easily apply counterfactual reasoning to social rules, but find it difficult to do so with logical ones; finally, (d) explain why, as pointed out by developmental psychological data, norm acquisition follows a stable ontogenetic pattern starting quite early in childhood ([24][25][26][27][28][29][30]).

In short, the intuition behind our normative architecture is twofold: on one hand, the emergence of norms is based upon a universal capacity to tell norms; on the other, this capacity is supported by a norm frame, an internal "model of a norm", which agents use as a processing instrument in norm recognition.

The emphasis laid on the innate and universal features of EMIL-A should not be mistaken, leading to think that no space is left to subjective variability. If norm recognition is a must, equally accomplished by a vast majority of agents, moral attitudes - i.e. the results of normative and moral experience accumulated during

lifetime that affect different normative procedures - are not. They are definitely subjective.

Furthermore, the reinforcement effects that occur on different EMIL-A procedures vary among agents. Personal experience, for example, impacts on norm salience. Analogously, the normative frame, being in constant interaction with the social environment and the other procedures, is liable to their influence. In these terms, a normative architecture is allowed to elegantly ignore the culture/nurture controversy.

4 Simulating norm emergence

Some simulation studies about the emergence of social norms have been carried out, for example Epstein and colleagues' study of the emergence of social norms ([31]), and Sen and Airiau's study of the emergence of a precedence rule in the traffic ([32]). In these studies, social norms are essentially seen as conventions, that is, behavioural conformities that do not imply explicit agreements among agents, and do emerge from their individual interests. Within this perspective, the function of norms is found in allowing participants in coordination games to choose one among equivalent alternative equilibriums. Agents repeatedly interact with other agents in social scenarios. Such interactions can be formulated as stage games with multiple equilibriums ([33]), which make coordination uncertain. Norms gradually emerge from interactional practice, essentially through mechanisms of imitation and social learning, establishing who should do what. So far, simulation-based studies have been applied to investigate which norm is chosen from a set of alternative equilibriums. In this framework agents are not provided with normative minds, but with strategic reasoning. No attention is paid to norm immergence, and therefore to the role of mental mechanisms in norm-emergence.

A rather different sort of question arises about the emergence of social norms when no alternative equilibriums are available for selection. This is a matter still not widely investigated and references are scanty if any ([34]). We propose that a possible answer to the puzzling questions posed above ought to be searched for by examining the interplay of communicated and observed behaviours, and the way these are interpreted and represented into the minds of the observers. If any new behaviour α is interpreted as obeying a norm, a new normative belief is generated into the agent's mind and a process of normative influence will be activated ([35]). We suggest that normative recognition represents a crucial requirement of norm emergence and innovation, as processes resulting from both agents' interpretations of one another's behaviours, and transmission of such interpretations to one another.

4.1 The Norm Recognition Module at work

Our Norm Recognizer (see [9][8] for a detailed description) consists of a long term memory, the normative board, and in a working memory, presented as a three layers

architecture. The normative board contains normative beliefs, ordered by salience. The difference in salience between normative beliefs and normative goals has the effect that some of these normative mental objects will be more active than others and they will interfere more frequently and with more strength with the general cognitive processes of the agent[2]. The working memory is a three layer architecture, where *social inputs* are elaborated. These inputs are represented on an ordered vector, consisting of four elements: the source (x); the type of input through which the message is presented (T)[3]; the addressee (y); the action transmitted (a). Agents observe or communicate social inputs. Once received the input from another agent, the agent will compute, thanks to its norm recognition module, the information in order to generate/update her normative beliefs.

Here follows a brief description of how this normative 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 to not answer when asked", is received, it will directly access at the second layer of the architecture, giving rise to a candidate normative belief "One must answer when asked", which will be temporally stored at the third layer. This will sharpen agents' attention: further messages with the same content, especially when observed as open behaviors, 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 corresponding normative behaviors observed, e.g. n% of the population), the candidate normative belief will be transformed in a new (real) normative belief, that will be stored in the normative board. The normative threshold can be reached in several ways: one way consists in observing a given number of agents performing the same action (alpha) prescribed by the candidate normative belief, e.g. agents answering when asked. If the agent receives no other occurences of the input action (alpha), after a fixed time *t*, the candidate normative belief will leave the working memory.

Aiming to decide which action to produce, the agent will search through the normative board: if more than one item is found out, the most salient norm will be

---

2 At the moment, the normative beliefs' salience can only increase, depending on how many instances of the same normative belief are stored in the Normative Board. This feature has the negative effect that some norms become highly salient, exerting an excessive interference with the decisional process of the agent. We are now improving the model, adding the possibility that, if the normative belief is inactive for a certain amount of time, its salience will decrease.

3 It can consist either in a *behaviour* (B), i.e. an action or reaction of an agent with regard to another agent or to the environment, or in a *communicated* message, transmitted through the following holders: assertions (A), i.e. generic sentences pointing to or describing states of the world; requests (R), i.e. requests of action made by another agent; deontics (D), partitioning situations between good/acceptable and bad/unacceptable; normative valuations (V), i.e. assertions about what it is right or wrong, correct or incorrect, appropriate or inappropriate (i.e. *it is correct to respect the queue*).

chosen.

5 The simulation model

The simulation model we designed is aimed to find out the sufficient (even if not necessary) conditions for existing norms to change. In particular, we want to show if a simple cultural or material constraint can facilitate norm innovation. We wonder if under such a condition, agents provided with a module for telling what a norm is can generate new (social) norms by forming new normative beliefs, irrespective of the most frequent actions. To see this, we imagined a simple case in which subpopulations are isolated in different contexts for a fixed period of time. The metaphor here is any physical catastrophe or political upheaval that divides one population into two separate communities. The recent European history has shown several examples of this phenomenon.

In our simulation model, the environment consists of four scenarios, in which the agents can produce three different kinds of actions. We define two context-specific actions for every scenario, and one action common to all scenarios. Therefore, we have nine actions. Suppose that the first context is a postal office, the second an information desk, the third our private apartment, and so on. In the first context the action *stand in the queue* is a context-specific action, whereas in the second a specific action could be *occupy a correct place in front of the desk*. A common action for all of the contexts could be, *answer when asked*. Each of our agents is provided with a personal agenda (i.e. a sequence of contexts randomly chosen), an individual and constant time of permanence in each scenario (when the time of permanence is expired, the agent moves to the next context) and a window of observation (i.e. a capacity for observing and interacting with a fixed number of agents) of the actions produced by other agents. Norm Recognizers are also provided with the three-layer architecture described above, necessary to analyze the received information, and a normative board in which the normative beliefs, once arisen, are stored. The agents can move across scenarios: once expired the time of permanence in one scenario, each agent moves to the subsequent scenario following her agenda. Such irregular flow (each agent has a different time of permanence and a different agenda) generates a complex behavior of the system, tick-after-tick producing a fuzzy definition of the scenarios, and tick-for-tick a fuzzy behavioral dynamics.

We have modeled two different kinds of environmental conditions. In the first set of simulations, agents can move through contexts (following their personal agenda and in accordance with the personal time of permanence). In the second set of simulations, from a fixed time t, agents are obliged to remain in the context they have reached, till the end of the simulation: in this case agents can explore the contexts exchanging messages with one another and observing others' behaviors. When they reach the last context at time *t*, they can interact with same-context agents till the end of the simulation. We hope this second setting allows us to show that the mere

statistical frequency is sufficient (but not necessary) to the agents' convergence on the common action.

At each tick, the Norm Recognizers (NRs), paired randomly, interact exchanging messages. These inputs are represented on an ordered vector, as said above. NRs produce different behaviors: if the normative board of an agent is empty (i.e. it contains no norms), the agent produces an action randomly chosen from the set of possible actions (for the context in question); in this case, also the modal by means of which the action is presented is chosen randomly. Vice versa, if the normative board contains some norms, the agent chooses the action corresponding to the most salient among these norms. In this case the action produced is presented with one of these modals: deontic (D), normative valuation (V) or behavior (B). This corresponds to the intuition that if an agent has a normative belief, there is a high propensity (in this chapter, this has been fixed to 90% of cases) for her to transmit it to other agents under strong modals (D or V) or open behavior (B). We run several simulations for different values of the threshold, testing the behaviors of the agents in the two different experimental conditions.

5.1 Results and Discussion

We briefly summarize the simulation scheme. The process begins by producing actions (and types of inputs) at random. The process is synchronic. The process is more and more complex runtime: agent i provides inputs to the agent who precedes her (*k=1*), issuing one action and one modal. Action choice is conditioned by the state of her normative board. When all of the agents have executed one simulation update, the whole process restarts at the next step.

First of all we present the results obtained when imposing the external barrier. Then, we present the results obtained when no barrier was imposed; finally we compare the former with the latter results.

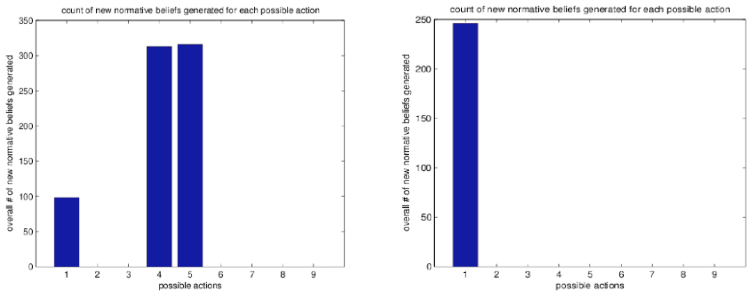

Fig. 1. (a-b). Overall number of new normative beliefs generated for each type of possible action - with (left) and without (right) external barrier

Figure 2(a) and Figure 2(b) show the trend of new normative beliefs generation runtime for a certain value of the norm threshold (threshold = 99), which is a good

implementation of our theory: each line represents the generation of new normative beliefs corresponding to an action (i.e. each line corresponds to the sum of different normative beliefs present in all of the agents). To be noted, a normative belief is not necessarily universally shared in the population. However, norms are behaviors that spread thanks to the spreading of the corresponding normative belief. Therefore, they imply shared normative beliefs.

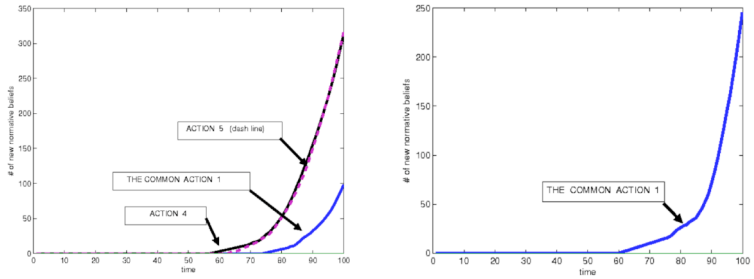

Fig. 2. (a-b). New normative beliefs generated runtime - with (left) and without (right) external barrier

Figures 3(a) and 3(b) are very similar (even if in the no-barrier variant - Figure 3(b), we find less regularity in the end of the dash line which represent the number of performed actions for the common action). In these plots, we cannot appreciate significant differences pointing to the normative beliefs acting on the effective behaviors: we cannot distinguish the clear effect corresponding to the agents' convergence on a specific norm (namely, we do not see that the dash line is significantly increasing).

Indeed, if we run longer simulations, we can appreciate the consequences of the results of our investigation: in Figures 4(a) and 4(b) we can observe two different (but related) effects: (i) more or less at the same time both in the barrier and no barrier condition, a convergence on the common action (dash line) is forming, much more significant in second case than in the first one; (ii) however, in the barrier condition, other lines of convergence are also emerging (increasing). If we observe Figure 5(a) and Figure 5(b) we can appreciate that in the first case (the case with barriers) we find a very very low convergence rate; but, in the second case (the case without barriers) we find a high convergence rate.

This corresponds to what is shown in in Figure 2(a) and Figure 2(b) on one hand, and Figure 1(a) and Figure 1(b) on the other: with external barrier, we can see that the higher overall number of new normative beliefs generated does not correspond to the common action (action 1) and the trend of new normative beliefs generated runtime shows the same results.

With no external barrier, instead, only normative beliefs concerning the common action (action 1) are generated.

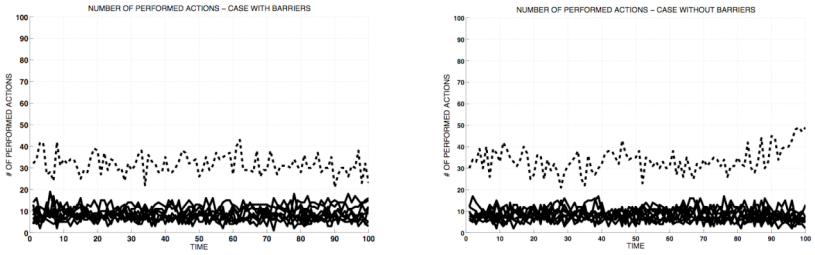

Fig. 3. (a-b). Actions performed by NRs - with (left) and without (right) external barrier. On axis X, the number of simulation ticks (100)is indicated and on axis Y the number of performed actions for each different type of action. The dash line corresponds to the action common to all scenarios

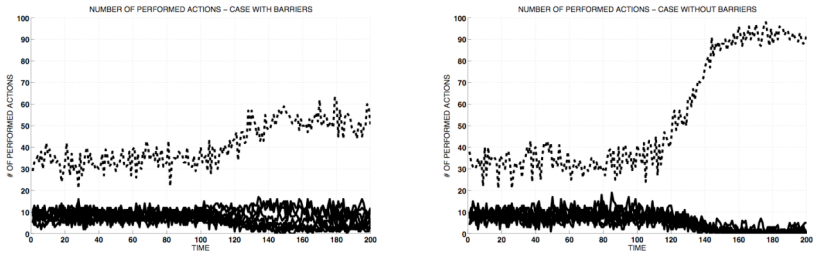

Fig. 4. (a-b). Actions performed by NRs - with (left) and without (right) external barrier. On axis X, the number of simulation ticks (200) is indicated and on axis Y the number of performed actions for each different type of action. The dash line corresponds to the action common to all scenarios.

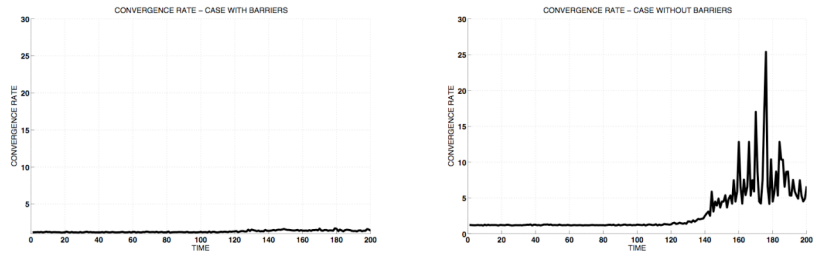

Fig. 5. (a-b). On axis X, the flow of time is shown; on axis Y the value of convergence rate in both cases with (left) and without (right) external barrier.

We have shown how our model allows new norms, which do not corresponding to the common action to emerge. Some rival norms now compete in the same social settings. Obviously, they will continue to compete, unless some further external event or change in the population (e.g. the barrier removal) will cause agents to start

migrating again. It would be interesting to observe how long the rival norms will survive after barrier removal, whether and when one will out-compete the others, and if so, which one. It should be said that, as we observe a latency time for a normative belief to give rise to a new normative behaviour, we also expect some time to elapse before a given behaviour disappears while and because the corresponding belief, decreasingly fed by observation and communication, starts to extinguish as well. We might call such a temporal discrepancy *inertia* of the norm. Both latency and inertia are determined by the twofold nature of the norm, mental and behavioural, which reinforce each other, thus preserving agents' autonomy: external barriers do modify agents' behaviours, but only through their minds.

More than emergence, our simulation shows a norm innovation process; in fact, Figure 3(a) shows that, starting around tick=60, two normative beliefs appear in the normative boards and the overall number of these two new normative beliefs generated is three times higher than the overall number of normative beliefs concerning the common action 1. We might say that, if stuck to their current location by external barriers, norm recognizers resist the effect of majority and do not converge on one equilibrium only. Rather, they will form as many normative beliefs as there were competing beliefs on the verge of overcoming the normative threshold before the agents had been stuck to their locations.

No such effect is expected among agents whose behavior depends only from the observation of others. In sum, is statistical frequency sufficient for a norm to emerge? Beside action 1, common to the four contexts, other norms seem to emerge in our simulation. Normative agents can recognize a norm; infer the existence of a norm by its occurrences in open behavior under certain conditions (see the critical role of previous deontics); and finally spread a normative belief to other agents. Future studies are meant to investigate on the effect of barrier removal and the inertia of normative beliefs.

6 Concluding Remarks

Does the theory presented answer the questions raised in the introduction? In principle, it allows the first set of questions to be answered, as we present a normative agent architecture, and show by means of simulation its value added with regard to concurrent, simpler, but less efficient agent models. But what about the former two sets of questions, concerning the link between different types of norms and their dynamics in society? We believe the view we have proposed accounts for both these questions in a rather innovative way.

As to integration, the paper presents a view of norms that discerns normal from normative behaviour at the same time filling the gap between legal and social norms. The solution we have proposed consists of agents' capacity to recognize a subset of the communicative or behavioural inputs they transmit to one another as normative, and autonomously decide to convert them into normative outputs.

As to norm dynamics, not only the spread of behavioural regularities and good social conduct are accounted for – what is allowed also by simpler models – but also what Ullmann-Margalit called prescribed behaviour ([36]), accounting for the mandatory character attributable to any norm, including conventions, as Margaret Gilbert acutely pointed out.

Did we answer the questions posed in a conclusive manner? Did we answer all of the interesting questions that one might pose with regard to norms? Of course, not. In particular, we have not taken into sufficient account the processes leading to incorporate norms into action schemata ([37]), or those leading to an automated normative will being formed (which Josh Epstein calls thoughtless compliance), or factors and processes leading to several different forms of norm internalization, from internalized norms to automated normative actions. In future works, we intend to integrate these ideas into our agent architecture in order to explore the effect of norm internalization on norm compliance and other relevant indicators of social order.

Acknowledgments. This work was supported by the EMIL project (IST-033841), funded by the Future and Emerging Technologies program of the European Commission, in the frame-work of the initiative Simulating Emergent Properties in Complex Systems.

References

1. Conte, R., and Castelfranchi, C. Cognitive and social action. University College of London Press, London (1995).
2. Conte, R., and Castelfranchi, C. From conventions to prescriptions. Towards a unified theory of norms. *AI&Law* 7: 323-340 (1999).
3. Conte, R., and Castelfranchi, C. The Mental Path of Norms. *Ratio Juris* 19 (4): 501 – 517 (2006).
4. Conte, R., and Castelfranchi, C. The mental path of norms. Ratio Juris, 19(4):501–517 (2006).
5. Castelfranchi, C. Simulating with cognitive agents: The importance of cognitive emergence. Multi-Agent Systems and Agent-Based Simulation, Springer, Berlin (1998).
6. Andrighetto, G., Campennì, M, Conte, R., Paolucci, M. On the Immergence of Norms: a Normative Agent Architecture. In *Proceedings of AAAI Symposium, Social and Organizational Aspects of Intelligence, 8-11 November 2007, Washington DC (2007).*
7. Conte, R., Andrighetto, G., Campenni', M., and Paolucci, M. Emergent and immergent effects in complex social systems. In Proceedings of AAAI Symposium, Social and Organizational Aspects of Intelligence, Washington DC (2007).

8. Campennì, M., Andrighetto, G., Cecconi, F., Conte, R. Normal = Normative? The role of intelligent agents in norm innovation, Mind & Society, 2009, 10.1007/S11299-009-0063-4 (2009).
9. Andrighetto, G., Campennì, M., Cecconi, F., Conte, R. The Complex Loop of Norm Emergence: a Simulation Model, in K. Takadama, C. C. Revilla, G. Deffuant (Eds.) The Second World Congress on Social Simulation, Springer-Verlag LNAI. Aspects of Intelligence Washington DC, 2008. (Forthcoming).
10. Rao, A. S. and Georgeff, M. P. Social plans: Preliminary report. In Werner, E. and Demazeau, Y., editors, *Decentralized AI 3 - Proceedings of the Third European Workshop on Modelling Autonomous Agents and Multi-Agent Worlds (MAAMAW-91)*, pages 57-76. Elsevier Science Publishers B.V.: Amsterdam, The Netherlands (1992).
11. Broersen, J., Dastani, M., Hulstijn, J., Huang, Z., and van der Torre, L. The BOID architecture. Conflicts between beliefs, obligations, intentions and desires. In Proceedings of the fifth international conference on Autonomous Agents, Montreal, Quebec, Canada, pp 9 – 16 (2001).
12. Wright, G. H. von. *Norm and Action. A Logical Inquiry*. Routledge and Kegan Paul, London (1963).
13. Kelsen, H. General Theory of Norms. Hardcover (1979).
14. Conte, R. Rational, goal governed agents. Encyclopedia of Complexity and Systems Science, Springer (2009).
15. Castelfranchi, C. Prescribed mental attitudes in goal-adoption and norm adoption. Artif. Intell. and Law, 7(1):37–50 (1999).
16. Wason, P. and Johnson-Laird, P. *Psychology of Reasoning: Structure and Content*. Harvard University Press, Cambridge, MA (1972).
17. Schank, R. C., and R. P. Abelson. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates (1977).
18. Fiske, S. T. and Taylor, S. E. *Social cognition* (2nd edn.). New York: McGraw Hill (1991).
19. Barsalou, L. W. Perceptual symbol systems. *Behavioral and Brain Sciences* 22: 577-660 1999.
20. Markus, H., and Zajonc, R. B. The cognitive perspective in social psychology. In G. Lindzey & E. Aronson (Eds.), Handbook of social psychology, pp. 137-229, 3rd Edition. New York: Random House (1985).
21. Sripada, C., and Stich, S. A Framework for the Psychology of Norms. In P. Carruthers, S. Laurence and S. Stich, eds., *The Innate Mind: Culture and Cognition*, 280-301, Oxford University Press (2006).
22. Cosmides, L., & Tooby, J. Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.). *The adapted mind*, New York: Oxford University Press (1992).
23. Cosmides, L., & Tooby, J. Can evolutionary psychology assist logicians? A reply to Mallon. In W. Sinnott-Armstrong (Ed.), *Moral psychology*. (pp. 131-136) Cambridge, MA: MIT Press (2008).

24. Bandura, A. Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50*, 248-287 (1991).
25. Nucci, L. P. *Education in the Moral Domain*. Cambridge University Press (2001).
26. Cummins, D. D. Evidence for deontic reasoning in 3- and 4-year olds. *Memory and Cognition* 24(6): 823-829 (1996).
27. Piaget, J. *The moral judgment of the child*. The Free Press: New York (1965).
28. Kohlberg, L. Justice and reversibility. In Kohlberg L, *Essays on Moral Development, vol. 1*. Harper and Row (1981).
29. Kohlberg, L., and Turiel, E. Moral development and moral education. In G. Lesser, ed. Psychology and educational practice. Scott Foresman (1971).
30. Shweder, R., Mahapatra, M. and Miller, J. Culture and moral development. In J. Kagan & S. Lamb (eds.), *The Emergence of Morality in Young Children*. The University of Chicago Press (1987).
31. Epstein, J. Generative Social Science. Studies in Agent-Based Computational Modeling. Princeton University Press, Princeton-New York (2006).
32. Sen, S. and Airiau, S. Emergence of norms through social learning. In Proceedings of the Twentieth International Joint Conference on AAAI (2007).
33. Myerson, R. B. Game Theory: Analysis of Conflict. Harvard University Press (1991).
34. Posner, R. and Rasmusen, E. Creating and enforcing norms, with special reference to sanctions. INT REV LAW ECON, pages 369–382 (1999).
35. Conte, R., and Dignum F. From Social Monitoring to Normative Influence. *JASSS* 4 (2) (2001).
36. Ullmann-Margalit, E. The Emergence of Norms. Clarendon Press, Oxford (1977).
37. Bicchieri, C. The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press, New York (2006).