The Institutional Stance in Agent-based Simulations

Giovanni Sileno, Alexander Boer and Tom Van Engers

Leibniz Center for Law, University of Amsterdam, Amsterdam, The Netherlands

Abstract: This paper presents a multi-agent framework intended to animate scenarios of compliance and non-compliance in a normative system. With the purpose of describing social human behaviour, we choose to reduce social complexity by creating models of the involved agents starting from stories, and completing them with background theories derived from common-sense and expert knowledge. For this reason, we explore how an institutional perspective can be taken into account in a computational framework. Roles, institutions and rules become components of the agent architecture. The social intelligence of the agent is distributed to several cognitive modules, performing the institutional thinking, whose outcomes are coordinated in the main decision-making cycle. The institutional logic is analyzed from a general simulation perspective, and a concrete possible choice is presented, drawn from fundamental legal concepts. As a concrete result, a preliminary implementation of the framework has been developed with *Jason*.

1 INTRODUCTION

In the last decades, software engineering has been moving away from machine-oriented views of programming towards concepts and abstractions more closely reflecting the way in which humans conceive the world. Multi-agent systems have been introduced to develop complex services, typically strongly characterized by interactions between entities, but, differently from other distributed systems, they often integrate concepts derived from philosophy or psychology, like the Belief-Desire-Intention paradigm. Furthermore, multi-agent systems are successfully used for simulations in economics, sociology, and biology[1] (in that case the term *ABM - Agent-based modelling* is preferred). However, although since a few years there has been an active research interest on normative multi-agent systems, there is no equivalent development about ABM simulations from a legal perspective. Although there are interesting overlaps with these studies there is an intrinsic, and important, difference of perspective. With the first approach, *conception*, we design and create artefact systems, which are largely under our control, trusted, and highly predictable (for example, electronic institutions). With the second approach, *interpretation* or *discovery*, we interpret the relevant events against a background of

systems, in our case mostly of human origin, some of which we trust to work in certain ways, and some others we only guess at. The second is the arena where ABM actually plays a role, and where this study should be placed.

2 AGENTS, INSTITUTIONS, ROLES

A (soft) Evolution. In the computational world, all starts (and ends) with imperative commands in the form of *instructions*. Instructions can be grouped in *procedures*. Data passes from numeric to string, unified in arrays, structures, and finally instantiated in objects, elements of classes defined by data structures and methods handling that data. *Objects* do not perform any action independently, they can only be used[2]. An *agent* is instead a proactive entity (*agere* in Latin means to do, to achieve, to lead). The basic type of agents is the *behavioural agent*: it has apparent autonomy, executes a *plan* composed by *actions*. The next step in this evolution is brought by *intentional agents*, that show autonomy, proactiveness, reactiveness and social ability. These four character-

---

[1]See for example in (Batten, 2000; Duffy, 2006).

[2]A practical definition in *object-oriented programming* may be different.

istics are the references of a very important human capability: *mentalization* (Fonagy and Target, 1997). Humans are able to *mentalize*, i.e. to create an internal representation of an external agent, in a way that its behaviour can be predicted and, according to the intentional approach, explained in terms of *attitudes*, for example using concepts like *beliefs*, *desires* and *intentions*[3]. This is an innate exercise of modelling for humans, and its importance comes from the fact of being commonly used as representation of reality. As a matter of fact, humans tend to mentalize not only individuals, but also communities, cultures, organizations, nations.

Instructions, Habits, Rules. Every human individual or group, having a goal, sets the more adequate means, or better, those considered as the more adequate means to achieve it. The means/end relationship creates generally an *instruction* of conduct of that type: *"If you want to reach the goal A, you have to do the action B"*. All these instructions could be very different in aim, content, scope of validity, or subjects involved. Moreover not all the instructions are established in the previous form (*to do*), but also in a qualification or state discriminant form (*to be*). If the former type requires the capability of executing tasks and in a larger sense of *planning*, the latter subsumes the existence of another process: the *situation (or state) recognition*. Anyway, all instructions have something in common: they are propositions that aim to influence the behaviour of individuals and/or groups, leading their actions towards certain objectives more than others. In a social agglomerate, similar instructions could be part of common knowledge, or cultural conventions, social habits and social rules. Only effective legal norms become part of the social rules.

Choice of Conduct. *Recognition* of the present state, and then *planning* tasks towards a *goal* are parts of a typical *problem solving* cycle (Breuker, 1994), that could be seen as the main core of the agent activity. The choice of conduct is then a sort of preference order between possible plans of actions, considering eventually associated economic costs or some potential side-effects. But how are norms related to the choice of conduct? (Neumann, 2010) has found in the normative multi-agent literature two main philosophical approaches: *deontic* and *consequentialistic*. The first considers norms as cognitive objects. A deontic conceptualization of norms typically emphasises that norms are in itself a reason for action. On the other
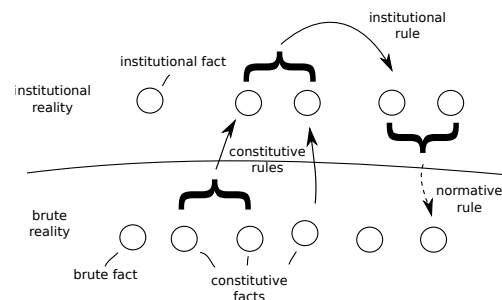

Figure 1: Institutional perspective.

side, with a consequentialistic approach, agents consider the possibility of breaking the norm as one of the alternative choices. Stated differently, the latter treats norms as behaviour regularities, and, for this, they may be described for example with game theoretic models[4]. The two different perspectives result in two potential attitudes towards norms; our target ABM framework should integrate both of them.

Institutions. In (Searle, 1969) an interesting point is traced about the difference between normative (or regulative) and constitutive rules. The former regulates existing forms of behaviour. For example, the rules of polite table behaviour have been introduced later than the eating activity. Eating existed and exists anyway, with or without these rules. But, at the contrary, the rules of playing chess have created the possibility of playing chess. Searle extended these examples. "The institutions of marriage, money and promising are like the institutions of baseball and chess in that they are systems of such constitutive rules or conventions."

Generalizing, we can say that an institution is an intentional social collective entity (Boer, 2009), defined by certain rules and some institutional facts. It is collective and intentional, simply because a group of people recognizes and intends its existence. This concept of institution unifies games, social informal norms and legal norms[5].

---

[3]The *intentional stance* has been introduced by Dennett (Dennett, 1987).

[4]From a game theory point of view, a social rule eliminates certain strategies for each of social convention the agents, and thus induces a *sub-game*. "For a given agent, a social law presents a trade-off; he suffers from loss of freedom, but can benefit from the fact that others lose some freedom." (Shoham and Leyton-brown, 2009)

[5]According to the institutional perspective, law is an institution whose purpose is to create normative order via formalization (MacCormick, 1998). Legal facts are dependent on legal norms and observed through legal norms. Law's 'truth' is different from reality, psychological or social truth. "In court proceedings, the 'truth' of the facts is ultimately determined not by criteria employed in empirical sciences but by those provided by procedural and substan-

A general institutional perspective has been visualized on Fig. 1. Two realities coexist: a brute reality and an institutional reality. Filtered by the *constitutive rules* associated to the institution, only few of all real facts determine an *institutional fact*. Stated differently, these real facts are events triggering a change in the institutional reality. They will be called *constitutive facts*, and typically are created through a *constituting act*. Institutional rules create new institutional facts from existing institutional facts. Finally, normative rules convert institutional facts into instructions of behaviour, and if the choice of conduct of the agent goes accordingly, they perform an actual normative action.

Institution as Agents. As pointed out in (Boer and Van Engers, 2011) externalization of rules and social organization is philosophically defensible, and often an efficient solution from a system engineering point of view. However, for ABM simulation purposes, this is not a natural choice. During a theory construction session with legal experts, it is more natural to think of the agent as provided with an internalized artefact, of arbitrary complexity, embedding the social factor. Following the terminology used in (Chu, 2011), we consider that there is no institution "*an sich*" in the brute reality. It exists only in the form of conceptualization, internal to each agent. The collective flavour of this type of entity is given by a similar representation in each individual of the community, or, better, may be regarded as an instance of emergence. At a macro-level, institutions arise, evolve and function as patterns of social self-organization, which goes beyond the conscious intentions of the individual humans involved, but still, it is generated by a micro-level, individual behaviour.

From the agent's point of view, to be complying with an institution means to have a representation of the present institutional state (rules and facts) and behave accordingly with the normative rules generated by reasoning with it. This is a specific form of the equation $Thinking = Knowledge + Reasoning$ (Kowalski, 2010). This kind of modularization unveils an architectural possibility: to consider the institution itself as a cognitive agent, dividing the model in separate but interconnected sub-models.

Scenarios and Roles. Humans continuously learn about the social institutions around them. The law tells us that a sale involves a buyer and a seller, that it deploys an offer and an acceptance, and that it leads

---

tive legal norms" (Tuori, 2006). This observation could be extended also to informal institutions.
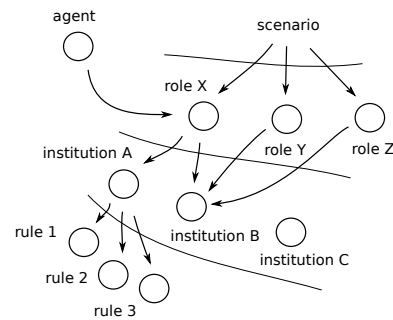

Figure 2: Architectural components.

to an obligation on the buyer to pay, and an obligation on the seller to deliver. It does not entail that the seller and buyer feel harmed if the other party does not meet their obligations, it does not tell what they will do if that happens, it does not tell when an offer will be acceptable, but we do have expectations about these things. In fact, analyzing a certain context, we provide a plausible (at least, according to us) intentional model of the involved agents, using our knowledge of the domain, analogies with other experiences and our mentalization capability.

In doing this, we use prototypes of *normal* behaviours (on which we will plausibly base our own behaviour, at the place of that agent), and also prototypes of the behaviours of non complying, *faulty* agents. Almost without noticing, we are not talking any more about individual agents. Identifying the contextual common patterns between individual cases (and their rationality), we are writing down typical *scenarios* (as patterns of social interaction) with the typical (social) *roles* that agents play in them. Such roles are associated to certain beliefs, plans (resulting in actual actions) and goals. Nothing forbids an agent to play simultaneously several roles (Fig. 2) and viceversa.

3 INSTITUTIONAL THINKING

Constitutive and institutional rules could be expressed in the form of normative conditionals (Sartor, 2006): $if\ CONDITIONS\ then^n\ CONCLUSION$. Conditions are constitutive and institutional facts. The conclusion is a new institutional fact. Institutional facts concern institutional entities or jural relations. Every new institutional fact typically draws a new configuration of the jural relations between the agent entities existing in that institution. This new configuration is a result of the *modus ponens* or forward reasoning applied through some deontic rules. We have taken as a starting point the fundamental legal concepts elaborated in

(Hohfeld, 1917), completed and further developed in (Sartor, 2006), but another choice could be possible as well.
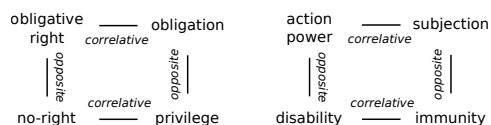

Figure 3: The two Hohfeldian squares: directed obligations and power concepts.

Directed Obligations and Permissions. Simplifying, the first Hohfeldian square defines the directed obligations between two actors, and it introduces four different deontic concepts: *obligative-right*, *obligation*, *privilege* and *no-right* (Fig. 3). An obligative-right, strictly defined, is one's enforceable claim against another (progressive form). Privilege is one's freedom from the claim of another one (regressive form). Like the cited author, we added to these concepts the *permissive-right*, different to obligative-right. This is in fact not completely reducible to the idea of a directed obligation.

Power. The second Hohfeldian square refers to the concept of power, with four new concepts: *action-power*, *subjection*, *disability*, *immunity* (Fig. 3). Action-power consists in the agent's power to determine a normative position involving a certain agent performing a certain action. On the contrary, disability defines the impossibility (*no-power*) to achieve a certain normative position.

Following the analysis found in (Vatiero, 2010), we extend their use also to not strictly legal aspects. In the case of the buyer and seller, if the buyer is completely dependent to the seller because he is the only one in possession of the good he wishes, he is actually in a subjection position towards the seller for his pricing choices. On the other hand, the seller is in a power position and can eventually raise the price. For the buyer, the only way to avoid this subjection is to stop wishing that good (which is not always possible), or finding another seller with a better price.

Generalizing this example, an action (in the form of an investment) of the agent may bring about a change of his institutional position. At the same time, contextual factors may determine an endogenous enforcement of power: the fact that the buyer needs that good and that the seller is the only one to sell determines the power positions between the two agents, and this is not influenced by the agents' consciousness of these positions. However, renewing the Bacon's *knowledge is power*, an agent that is able to recognize

the contextual state, and knows how to change it accordingly to his own interests, has the action-power to modify the power positions in the institution.

Coordination and Competition Dimensions. The agent behaviour associated to an institution could be generally ascribed along two main axis: *(social) co-ordination* and *(economic) competition*. Simplifying, the first covers the rules of the "game", the second the knowledge about how to "play the game" within the rules. The rules of *fair competition* are particularly meaningful. They are introduced in order to avoid that one player attains a position too strong compared to the others, limiting or blocking *de facto* their possibilities of action and resulting in a failure or degeneration of the institution.

Practical Normative Indications. In order to have a successful normative action, the jural relations resulting from the institution should have some consequence on the agent's behaviour. Focusing on the decision-making cycle, only four practical normative indications are important for the agent: you can[6] *[7], you can not *, you must *, you do not have to *.

In general, * may be related to actions (to do), or states (to be), or results (to bring about). The *state* subsumes a state recognition, i.e. the possibility that a certain group of perceived/believed conditions counts as a certain state. For this reason, defining a fact as a reified statement about something, a state/situation is a second order fact, requiring a qualification. The *bring about* requires a plan toward the objective/result, and a result can be both a fact or a state.
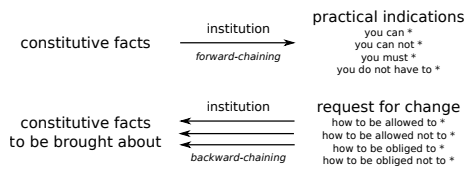
Figure 4: Forward and backward reasoning in institution.

Institutional Backward Reasoning. The process analyzed until now, from constitutive facts to practical indications, is a forward-chaining procedure (*condition to conclusions*). It is interesting to consider also the other way around, the backward-chaining (*conclusion, if condition*): which conditions have to be brought about to reach a different institutional configuration? More answers might be possible to the same question, in the form of alternative sub-goals aiming to the same objective (Fig. 4).

If the logic describing rules and jural relations is consistent, forward and backward chaining relations can be automatically inferred one from the other. However, here we emphasize the possibility to simulate in such a framework also a wrong conceptualization of the institution, conflicting rules, faulty logic, using for example a non-monotonic approach and handling separately the forward and backward relations[8].

Normative Plans. Put in a simple form, in order to be complying, an agent should: (a) check, before performing an action, if he has the permission to do that, and if not, evaluate to drop the plan involving that action, (b) add as goal the performance of any new obligation. On the other side, in order to be socially reactive in a context in which non-complying behaviours are possible, an agent should: (c) be able to recognize not complying events or situations[9], (d) have a maintenance plan with a monitoring purpose, (e) react to the occurrence of not complying events.

---

[6]In this context *can* has only the meaning of *being allowed to* (and not, for example, of *being able to*).

[7]Here we are considering * defined always in a positive form, but it is easy to demonstrate that the negative forms can be reduced to one of these. Furthermore, all four positions could actually be expressed using only one functor, nevertheless we prefer to use the proposed expressions, as they are more pertinent to focus on behaviour.

[8]Starting from the famous Wason's selection task, several studies have been conducted in behavioural psychology with the purpose of assessing the heuristics involved in human reasoning. Part of these results, relevant for our scopes, are described in (Cosmides and Tooby, 2008).

[9]Using the terms previously defined, an event is a fact, observed or acknowledged by the agent, and a situation is a state, in the sense of fact of facts.

Problem Solving Cycle of the Agent. From the point of view of the modeller, it is worth to consider the possibility to associate different attitudes towards norm to each institutional component. Following the consequentialistic approach the agent will consider plans containing violations of norms. Following the deontic approach he will abolish any plan that includes non-complying actions or results. In both cases, some kind of meta-reasoning about plans has to be implemented. For this reason, this investigation would be part of the more general conflict resolution problem for *norm vs desire* conflicts, to be analyzed altogether with *plan vs plan* (when multiple alternative plans are possible to reach the same goal), *desire vs desire* conflicts (caused for example by conflicting goals between simultaneous roles), *norm vs norm* (between institutions) and *rule vs rule* conflicts (for an internal institutional conflict)[10]. From a general point of view, we are not aiming for autonomous agents that create their own plans from first principles, but that evaluate and choose between existing plans according a given rationality.

4 PRELIMINARY IMPLEMENTATION

An implementation of the framework has been developed for *Jason*, a MAS platform based on an extension of the *AgentSpeak(L)* programming language[11] (Bordini et al., 2007). The source code is available on our site[12] and consists in different modules handling the creation of institutions, the institutional logic, the institutional communication and a coordination solution that follows a deontic approach towards normative indications. At the moment, two examples are included, not presented here in details for space constraints. The first one describes a mythological story of non-compliance (Achilles avoiding the Trojan war): an example of how a story could be animated, starting from a narrative, interpreting it, and adding background theories derived from commonsense knowledge. The second one is an institutional model of the sale process, as described in common

---

[10]Many of these conflicts are consequence of the distribution of the the institutional thinking to multiple entities, thus requiring adequate coordination functions. However, this is also a good representation of our daily experience about the choice of conduct, when we have to take choices between our social roles expectations (for example as members of a family, researchers, citizens, etc.).

[11]*Jason* is based on logic programming and implements a BDI architecture for cognitive autonomous agents.

[12]*http://justinian.leibnizcenter.org/inst_jason*

law/the Dutch civil code, backing several buyer/seller scenarios.

5 CONCLUSIONS AND FURTHER DEVELOPMENTS

The intent underlying our research is to (re)connect normative (including legal) reasoning with other forms of reasoning. Particularly, we are interested in the role that (legal) norms play in social structures and how norms influence human behavior in those structures.

In the current approach, typical strategy decision problems for a given game do not take explicitly into account the possibility of the player to behave avoiding a rule, or forcing the interpretation of the rule toward its interest, if the regulator (consciously or not) left some ambiguity. The second case is not so common in games, but in the case of legal order, *lacunae* of law are practically unavoidable and within limits, desirable. This is part in fact of the human collective adaptation and social reasoning capabilities. In that way, humans question rules, both directly or in an involuntary way (for example, in the case of lack of knowledge or misunderstanding of the rule) and determine with their actions if social rules are successful or not in their normative intentions. Furthermore, humans do not *play* ever a single "*game*" at once. In a broader sense, humans are always players of several games simultaneously, or to put it differently, they are agents concerned in the same moment by many different institutions, sometimes conflicting, created by habits, social rules and legal order.

In the present paper we propose a framework that aims to take explicitly all of this into account. Our objective is a partial alignment of the representations of law with actual social structures and existing implementations of law. Descriptions of those are present for example in legal narratives, in the form of court decisions or anecdotes by legal experts, where a constructed theory, at least partially, is explicitly stated. Thus, using our framework, models of agents or roles involved in a social scenario could be animated, outlined from a story, enriched with knowledge from experts and/or referring to the sources of regulation, with the possibility of integrating game-theoretic behavioural theories. As operative result, such a simulation would furnish a support to understand the social (institutional) dynamics: validating the domain of conceptualization of the experts, making predictions, suggesting improvements to regulations.

Along with this paper, a preliminary implementation has been developed, using an existing multi-agent

system platform. Although successful, this experience showed the necessity of creating (or extending) a platform with an explicit ABM approach in order to attain a full computational deployment of the proposed framework. This is one of the directions of our future research.

REFERENCES

Batten, D. F. (2000). *Discovering Artificial Economics*. Westview Press.

Boer, A. (2009). *Legal Theory, Sources of Law and the Semantic Web*. IOS Press.

Boer, A. and Van Engers, T. (2011). An Agent-based Legal Knowledge Acquisition Methodology for Agile Public Administration. *ICAIL 2011: The Thirteenth International Conference on Artificial Intelligence and Law*, June.

Bordini, R. H., Hübner, J. F., and Wooldridge, M. (2007). *Programming multi-agent systems in AgentSpeak using Jason*. John Wiley & Sons Ltd.

Breuker, J. (1994). Components of problem solving and types of problems. *A Future for Knowledge Acquisition*, 867:118–136.

Chu, D. (2011). Complexity: against systems. *Theory in biosciences*, 130(3):229–45.

Cosmides, L. and Tooby, J. (2008). Can a General Deontic Logic Capture the Facts of Human Moral Reasoning? How the Mind Interprets Social Exchange Rules and Detects Cheaters. In Sinnott-Armstrong, W., editor, *Moral psychology*, pages 53–119. MIT Press, Cambridge.

Dennett, D. C. (1987). *The Intentional Stance*. MIT Press, Cambridge, Massachusetts, 7th edition.

Duffy, J. (2006). Agent-based models and human subject experiments. *Handbook of computational economics*, 2(05):949–1011.

Fonagy, P. and Target, M. (1997). Attachment and reflective function: their role in self-organization. *Development and psychopathology*, 9(4):679–700.

Hohfeld, W. N. (1917). Fundamental legal conceptions as applied in judicial reasoning. *The Yale Law Journal*, 26(8):710–770.

Kowalski, R. A. (2010). *Computational Logic and Human Thinking: How to be Artificially Intelligent*. Number November. Cambridge University Press.

MacCormick, N. (1998). Norms, institutions, and institutional facts. *Law and Philosophy*, 17(3):301–345.

Neumann, M. (2010). A classification of normative architectures. *Simulating Interacting Agents and Social Phenomena*, 7:3–18.

Sartor, G. (2006). Fundamental legal concepts: A formal and teleological characterisation. *Artificial Intelligence and Law*, 14(1):101–142.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Shoham, Y. and Leyton-brown, K. (2009). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge Univ Pr.

Tuori, K. (2006). Self-description and external description of the law. *No Foundations*, 2:27–43.

Vatiero, M. (2010). From W. N. Hohfeld to J. R. Commons, and Beyond? *Journal of Economics*, 69(2).