Informatica e diritto, XXXIX annata, Vol. XXII, 2013, n. 1, pp. 223-241

An Interdisciplinary Approach to Multi-agent Systems: Bridging the Gap between Law and Computer Science

MIGLE LAUKYTE*

1. INTRODUCTION

Research in Multi-agent Systems - MASs has given rise to new issues in sociology, psychology, philosophy, and other social sciences, all the while providing new insights into some abiding issues. But legal science has not quite responded to these developments: the computational simulation of legally relevant social activities and phenomena is a research area that has yet to hit its stride. Why is that so? And what can be done to encourage the development of such simulation?

This paper attempts to answer these questions by developing two related ideas that, if brought to fruition, could change the current situation for the better: the first is the interdisciplinary idea of a boundary object; the second, that of an agent's autonomy.

As concerns the first idea, that of boundary objects, I argue that an important reason why the simulation of legal phenomena is not making as much headway as legal scholars and computer scientists would like is a certain language barrier that lawyers and computer scientists (in particular, software engineers) have to overcome if they are to achieve the sort of fluent communication needed to create a successful legal MAS. I do not speculate about the *causes* of this language barrier, but I do point out that one way in which it can be taken down is through an approach that – by bringing to bear the sociological concept of a boundary object, understood as an interactive object

* The Author is a Max Weber Fellow at the European University Institute (EUI), Florence (Italy). Her studies focus on different aspects of the intersection among science, society, and new technologies, and she is particularly interested in investigating the role of law in mediating between scientific innovation and societal needs.

lying at the boundary between different disciplines – makes it possible for the relative research communities to relate to one another and work together in a mutually beneficial way (here, in building a legal MAS).

The second idea I bring into play and develop is that of the autonomy which can be ascribed to an artificial agent within an MAS. I work out a *legal* concept of autonomy by setting out the conditions satisfying which an agent's action can be deemed autonomous, in effect identifying a threshold of autonomous action that would trigger legal consequences. I argue that artificial agents have yet to reach this threshold, and I discuss the legal considerations that would have to be taken into account in light of this future development.

With that done, I tie this idea of autonomy to the previous discussion on boundary objects by treating autonomy itself as a boundary object. In other words, I illustrate how, if we can spell out in a clear enough way what autonomous action means in the law, and what the legal approach to such action ought to be, then we will also have a roadmap we can rely on in developing autonomous artificial agents and building MASs that would be more effective at simulating or reproducing social interactions in areas of activity which fall within the purview of the law. Autonomy so conceived would count as a boundary object by virtue of its lying at the intersection of law and computer science, but what is even more important is that the two broad communities (lawyers, jurists, and legal scholars on the one hand, computer scientists on the other) would be able to share ideas and solutions in working toward MASs "staffed" by artificial agents whose autonomy makes them at once practically intelligent (in replacing human agents) and legally cognizant.

2. Law and the Development of MASs

I take a broad view of MASs as any computational system made of artificial agents that interact and communicate with one another and with their environment, and where the system's overall behavior is more important than that of its single agents[1], and I take an equally broad view of artificial agents as any computational entity that has a role and can "do something" within such a system. An important application in this regard is the use of MASs to simulate or actually *do* the work that humans do in coordinative

[1] A. Omicini, *Challenges and Research Directions in Agent-oriented Software Engineering. Autonomous Agents and Multi-agent Systems*, in "Challenges for Agent-based Computing", Special issue, Vol. 9, 2004, n. 3, pp. 253-283.

enterprises, meaning any task requiring some kind of coordinated action or relational interactivity: this kind of simulation has been attempted, for example, in psychology[2], cognitive science[3], generative science[4], biology[5], economics[6], geography[7], political science[8], and transportation and logistics[9].

The common denominator to all these applications is that they all require different artificial agents to interact toward a common goal. Now, I would not go so far as to say that a legal system is driven by a common goal (other than that of peace and stability among equals[10]), but I would say that if anything can be identified as the object of law, that would have to be the interactive component of human endeavors, meaning that law arises wherever different people have to interact in such a way that one person's goals do not come into conflict with another's. Of course, this is a minimal, stripped

[2] S. BANDINI, S. MANZONI, G. VIZZARI, *A Multi-agent System for Remore Psychological Profiling with Role-Playing Games-based Tests*, in "Proceedings of the 2003 ACM Symposium on Applied Computing", New York, ACM, 2003, pp. 33-37.

[3] C. CASTELFRANCHI, *The Theory of Social Functions: Challenges for Computational Social Science and Multi-agent Learning*, in "Cognitive Systems Research", Vol. 2, 2001, n. 1, pp. 5-38; C. CASTELFRANCHI, F. DE ROSIS, R. FALCONE, S. PIZZUTILO, *Personality Traits and Social Attitudes in Multiagent Cooperation*, in "Applied Artificial Intelligence", Vol. 12, 1998, n. 7-8, pp. 649-675.

[4] J.M. EPSTEIN, *Generative Social Science: Studies in Agent-based Computational Modeling*, Princeton, Princeton University Press, 2011.

[5] C.E. SANSORES, F. REYES, H.F. GÓMEZ, J. PAVÓN, L.E. CALDERÓN-AGUILERA, *BioMASS: a Biological Multi-agent Simulation System*, in "Proceedings of the Federated Conference on Computer Science and Information Systems - FedCSIS", IEEE Explore Digital Library, 2011, pp. 675-682.

[6] A. SCHWAIGER, B. STAHMER, *SimMarket: Multiagent-based Customer Simulation and Decision Support for Category Management*, in Schillo M., Klusch M., Müller J., Tianfield H. (eds.), "Proceedings of the 1st German Conference on Multiagent Technologies (MATES 2003)", Lecture Notes in Artificial Intelligence, Berlin-Heidelberg, Springer Verlag, 2003, pp. 74-84.

[7] A.U. FRANK, S. BITTNER, M. RAUBAL, *Spatial and Cognitive Simulation with Multi-agent Systems*, in Montello D. (ed.), "Spatial Information Theory – Foundations of Geographic Information Science", Berlin-Heidelberg, Springer Verlag, 2001, pp. 124-139.

[8] T. YU, S.-H. CHEN, *Agent-based Modeling of the Prediction Markets for Political Elections*, in Villatoro D., Sabater-Mir J., Sichman J.S. (eds.), "Multi-agent Based Simulation", LNAI 7124, Berlin-Heidelberg, Springer Verlag, 2012, pp. 31-43.

[9] M.P. LUCK, MCBURNEY, O. SHEHORY, S. WILLMOTT, *Agent Technology: Computing as Interaction. A Roadmap for Agent Based Computing*, http://www.agentlink.org/roadmap/al3rm.pdf, 2005.

[10] See R.A. MANN, B.S. ROBERTS, *Essentials of Business Law and the Legal Environment*, Mason, South-Western Cengage Learning, 2010, p. 3.

down definition of law, and I do not so much pretend as to solve the abiding jurisprudential problem "What is law?" But at least this much can be agreed on, that wherever you have human interaction on a large scale (anything larger than face-to-face contact), there you have law, or the need for law, and wherever such a need arises, there also arises a need for rules. Rules are precisely the bread and butter of law, "the most pervasive and developed normative system"[11], the system we rely on to regulate human interaction (whether by design or by custom, or a combination thereof). So the question is: If law is so centrally concerned with regulating human interactions, and if MASs are so centrally concerned with simulating these interactions, why is there not much of an effort to design legal MASs, or MASs capable of simulating legally relevant social activities?

Such simulation could be put to use in different areas of law (civil, administrative, criminal, labour, family, etc.), drawing on the expertise of lawyers and legal philosophers alike, who in turn would find an interest in working out the ramifications of the activity in question. And, in fairness, some attempts have been made in working toward this goal[12]. But there is still much work to be done, and I speculate that the reason for this lies in part in a certain language barrier that is making it difficult for lawyers and computer scientists to work together. So, my idea is that if some common ground can be found – or a framework through which scholars and researchers who

[11] R. CONTE, R. FALCONE, *Introduction: Agents and Norms: How to Fill the Gap?*, in "Artificial Intelligence and Law", Vol. 7, 1999, p. 1.

[12] Examples are (*i*) the cadastral system described in A.U. FRANK, S. BITTNER, M. RAUBAL, cit., pp. 124-139, designed to simulate the transfer of rights in land, among other purposes, though the authors agree that in the future we will have to further investigate the way "simulations with multi-agent systems can help to explore how technical systems and legal requirements interact"; (*ii*) the crime-prevention model described in T. BOSSE, C. GERRITSEN, *A Model-based Reasoning Approach to Prevent Crime*, in Magnani L., Carnielli W., Pizzi C. (eds.), "Model-based Reasoning in Science and Technology Abduction, Logic, and Computational Discovery", Berlin-Heidelberg, Springer Verlag, 2010, pp. 159-177, which also makes policy recommendations based on the model; (*iii*) the development of MASs having the requisite of legal capacity to form binding contracts and carry out other legal transactions, see G. BOELLA, L. VAN DER TORRE, *Contracts as Legal Institutions in Organizations of Autonomous Agents*, in "Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems", Washington, IEEE Computer Society, Vol. 2, 2004, pp. 948-955; and, more generally, (*iv*) the development of normative MASs, or agent-based systems whose functioning is based on social norms, though it must be stressed that this line of research is still in its infancy, see A. ROTOLO, L. VAN DER TORRE, *Rules, Agents and Norms: Guidelines for Rule-based Normative Multi-agent System*, in Bassiliades N., Governatori G., Paschke A. (eds.), "Rule-based Reasoning, Programming, and Applications", Lecture Notes in Computer Science, Berlin-Heidelberg, Springer Verlag, 2011, pp. 52-66.

would not otherwise interact can do so – then we will have at least created the conditions for cooperation toward a common goal. It is here that the idea of a boundary object comes into play, for I view such objects as key to any cross-disciplinary endeavor. This will be the subject of the next Section, where I introduce boundary objects so as to illustrate how they apply to the case at hand.

3. The Interdisciplinary Idea of Boundary Objects

The idea of a boundary object was first introduced more than two decades ago, in 1989 by Susan Leigh Star and James Griesemer[13], the former a psychologist and sociologist interested in information technology and science, and the latter a philosopher interested in biology. What they meant by the term "boundary object" was any object that different communities of practice use and conceive in different ways while still recognizing the object for what it is. Objects of this kind are liable to take on a variety of meanings depending on who is using them and for what purposes. This may seem to make it more difficult for us to deal with such objects, but it is actually an advantage, for otherwise everything in the world would consist of "rigid" objects each of which boxes us into a single use of it according to its proper conception. A boundary object, by contrast, is at once stable and plastic, or malleable: it is solid enough at its core that the different communities using it will still know they are essentially dealing with the same object, but it is "soft" around the edges, and in this way we can fashion it into different "shapes" depending on what we are trying to do with it and how. If you are now thinking, in light of this description, that *anything*, even the apple you are about to eat, can be a boundary object, you are right, because the term to a good extent describes not so much a kind of object as an *attitude* we take toward objects (the attitude I hope to encourage). In fact the idea has undergone its own evolution over the years, expanding its range to include, among other things, processes and even human beings: the only "requirement is that they be able to span boundaries separating social worlds, so that those on either side can 'get behind' the boundary object and work together toward some goal"[14]. That is precisely the idea I want to stress, and

[13] S.L. Star, J. Griesemer, *Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39*, in "Social Studies of Science", Vol. 19, 1989, n. 3, pp. 387-420.

[14] T.F. Gieryn, *Boundaries of Science*, in Jasanoff S., Markle G.E., Petersen J.C., Pinch T. (eds.), "Handbook of Science and Technology Studies", London, Sage Publications, 1995, pp.

the corresponding collaborative attitude I would like to foster, in discussing the idea of a boundary object. But let us see how it all works by considering who the intended "users" of boundary objects are, the previously mentioned communities of practice.

3.1. Communities of Practice as Users of Boundary Objects

There are different conceptions of what a community of practice is[15], but perhaps the most comprehensive discussion of this idea is the one offered by Wenger[16], for whom a community of practice (or CP) is an "organic, spontaneous and informal"[17] group of people bound together by their expertise, interests, and goals, among other things. What makes them a community is (a) a shared interest in working on common problems and developing certain themes and (b) a shared repertoire of resources that they can draw on in that pursuit. The community would typically set out goals by involving its members in discussions and asking for contributions from them, but this does not mean that all members share the same goals. On the contrary, the landscape for action is wide, and so much latitude is afforded to the members to specifically frame their own goals within the broad outline. It follows that the goals framed by the community as a whole can be thought of as belonging at the same time to its members individually, precisely because each member is free to contribute to the community in an individual way. This introduces a dynamic nature of CP in that a CP evolves with and through its members, who do share common interest and a common store of knowledge, but they also shape that interest and carve that knowledge in specific ways, so that as a group they can engage in mutually reinforcing activities. As Wenger[18] puts it, "communities of practice create a dynamic form of continuity that preserves knowledge while keeping it current".

One can already see the basic analogy between this description of a CP and the previous description of a boundary object: just as a CP is bound by

414-415. I should note here that Star and Griesemer themselves used the term *social world*, though they did not mean it to be coextensive with the idea of a community of practice.

[15] For an overview see A. COX, *What Are Communities of Practice? A Comparative Review of Four Seminal Works*, in "Journal of Information Science", Vol. 31, 2005, n. 6, pp. 527-540.

[16] E. WENGER, *Situated Learning: Legitimate Peripheral Participation*, Cambridge, Cambridge University Press, 1991; E. WENGER, *Communities of Practice: Learning, Meaning and Identity*, Cambridge, Cambridge University Press, 1998.

[17] E. WENGER, *Communities of Practice: The Organizational Frontier*, in "Harvard Business Review", 2000, Jan.-Feb., p. 140.

[18] E. WENGER, *Communities of Practice: Learning, Meaning and Identity*, cit., p. 252.

a core interest that each member develops in individual ways, so a boundary object can be said to have a core shape (that makes it recognizable as such) whose contours can be fashioned in different ways. And just as virtually anything can be thought of as a boundary object, so virtually any group of people can be thought of as a CP: lawyers can form a CP just as moms can, for example, or even neighbors. Also, any person is likely to be in several CPs at once, at different levels of engagement, and for this reason no CP can be conceived as an isolate: CPs exist only as so many parts of the extensive, overarching social network. Wenger[19] expresses this idea by presenting CPs as "histories of articulation with the rest of the world", arguing that CPs "take the responsibility for the preservation of old competencies and the development of new ones, for the continued relevance of artifacts, stories, and routines, for the renewal of concepts and techniques, and for the fine tuning of enterprises to new circumstances"[20]. So let us consider a specific example of a CP, so that we can then see in practical terms how boundary objects fit into the discussion: let us take the community of artists. This community is made of people who share

- an interest in making art (expressing content through form and color) and
- a common vocabulary and general store of knowledge that they draw on in pursuing that activity.

The activity of making art is carried out in a dynamic way: it evolves, and furthermore, it does so by virtue of the mutually reinforcing engagement through which different artists go about their work. And it is here that boundary objects come into play: the vocabulary, materials, techniques, and ideas forming the common stock of resources that artists draw on in making art are themselves the boundary objects that enable artists to make art, or create, both as individuals and as a community.

How so? To see this, we need only to go back to the idea of a boundary object as something defined by a hard core enveloped in a soft shell. Suppose the object in question is art itself. We may not be able to define art, but at least we can recognize it when we see it. And that is precisely what boundary objects do: they enable the members of different CPs (artists, critics, consumers) to recognize them as objects of a certain kind (a work of art, a cubist painting, a watercolor), all the while making it possible for each CP to fashion and conceive the object in different ways, this (a) depending on what

[19] *Ivi*, p. 103.
[20] *Ivi*, p. 252.

a particular CP is trying to achieve, and (b) by sharing knowledge and ideas with others in the same community (with other artists, critics, or consumers of art).

As can be surmised from the example just made, boundary objects are something of a double-edged sword: they can promote collaboration among different communities, but they can just as easily be a source of dissension, precisely because each such object is amenable to a variety of uses and interpretations. In fact, each CP is liable to assert its own understanding of a given boundary object, rejecting all other conceptions of it as irrelevant or even wrongheaded. This is an inherent risk but not necessarily a disadvantage: as I am trying to argue, diversity of interpretation can be an asset, rather than a liability, because it can support a dynamic of mutual reinforcement among different CPs.

A final point is that the multiplex nature of boundary objects holds not only for abstract entities (like art, color, form, and realism) but also for more concrete ones, like color contrast, *chiaroscuro*, and even oil color. In other words, just as we can all recognize something as art even though we may disagree on its way of "capturing" that concept, so we can all recognize oil color as such while making different uses of it depending on what purposes we are driven by in that regard. For example, a painter and a manufacturer might have an equal interest in the permanence and stability of a given oil color but will be driven by different purposes, the former looking to achieve the color effect with which to express himself, the latter instead looking to market a product recognized for its quality. So, in summary, a boundary object does not just admit of different uses and corresponding interpretations, but would not even exist without a constellation of CPs revolving around it, that is, without different communities using it in different (possibly mutually empowering) ways.

3.2. Lawyers and Computer Scientists: Two CPs on Different Sides of the Same Boundary Object - The Artificial Agent

Let us see how the concepts just illustrated can be brought to bear on the idea of agents as boundary objects providing an interface between lawyers and computer scientists. The point of the discussion is, of course, to show how the two communities might be able to work closer together, to the degree that one would expect given the legal implications of what it means to use artificial agents to simulate human interaction in various areas of activity.

At first glance, the two CPs – lawyers on the one hand, computer scientists on the other – seem quite removed from one another, the former working in what might be termed a social science, or at least a human science, and the latter in an exact science, especially as concerns software engineering[21]. So let us see how two apparently unrelated CPs might be able to come together and collaborate in dealing with artificial agents (the building blocks of a MAS).

To begin with they need to identify a common interest, or a broad problem that they work on in tandem. This was hinted at earlier on in the discussion when it was pointed out that every MAS is an interactive system and that interaction – or rather, the *rules* by which interaction among artificial agents is regulated – is a central concern of law. This introduces at least two lines of research centered on the idea of a rule. The first of these can be captured in the question (*a*) How can artificial agents be designed to simulate the actual working of legal rules, or the legal effects of human action and interaction? This line of research can in turn be broken down into two strands, concerned with the simulation of either (*a.1*) acts *in* the law or (*a.2*) acts *of* the law, the former class comprising any "act that is intended to create, transfer, or extinguish a right and that is effective in law for that purpose; the exercise of a legal power (also termed *juristic act*, *act of the party*, *legal act*)", and the latter "the creation, extinction, or transfer of a right by the operation of the law itself, without any consent on the part of the persons concerned (also termed *legal act*)"[22]. And the second line of research can instead be encapsulated in the question (*b*) How are we to work out the legal implications of what artificial agents do *as artificial agents*, rather than as agents designed to simulate *human* action and interaction under the law?

And now, having identified a common interest between two otherwise unrelated CPs, we will have to identify a shared vocabulary. We do so by identifying the appropriate boundary objects, and two candidates immediately suggest themselves in this respect: the first is the concept of a rule (as just considered) and the second that of an agent. The former concept I will not take up here: it is a subject best discussed at some length in a dedicated investigation. The latter concept, that of an agent, is likewise a boundary object in that the two CPs clearly use the same term in different ways – the

[21] C.A.R. HOARE, *An Axiomatic Basis for Computer Programming*, in "Communications of ACM", Vol. 12, 1969, n. 10, pp. 576-583.

[22] S.v. "Act in the law" and "Act of the law" in B.A. GARNER (ed.), *Black's Law Dictionary*, 7th ed., St. Paul, West Group, 1999.

term *agent*, designating *artificial* agents for one CP, and someone acting on another's behalf for the other – but they also recognize that at some level they are dealing with the same thing. One might justifiably ask here how that is possible, considering the distance that separates the two CPs, but remember that a boundary object always has a solid core, and I submit that this solid core lies in the idea of autonomy, in that nothing can be described as an agent unless it is to some extent capable of autonomous action.

That is an idea to which we turn in the next Section. But in the meantime I should note that what applies to *agents* as boundary objects also applies to *artificial* agents as such objects. Here, too, we can see that software engineers and lawyers will approach artificial agents from different angles, the former with an interest in developing MASs designed to carry out complex, interactive tasks, the latter with an interest in working out the legal consequences that flow from such activity (research line *b* above). Software engineers accordingly treat artificial agents as abstract entities forming part of an overall interactive complex of elements, each driven by a goal functional to the broader goal for which an MAS is developed[23]. And lawyers, for their part, see artificial agents as fictions, that is, they create the legal fiction of an artificial agent as an *actual* agent (a person) so as to apply to that entity the rules that would apply to any other subject in a similar position under the law; or they see artificial agents as products, with a corresponding interest in product liability and the law that applies to products as works of authorship (copyright) or as inventions (patent law)[24]. But even here, where artificial agents are concerned, we can take the concept of an agent's autonomy as the core content through which different communities can relate in dealing with artificial agents[25]. And so an artificial agent can be understood as a boundary object having autonomy as a core property through which different CPs

[23] On artificial agents as abstractions, see M. WOOLDRIDGE, P. CIANCARINI, *Agent-Oriented Software Engineering: The State of the Art*, in Ciancarini P., Wooldridge M. (eds.), "Agent-Oriented Software Engineering", Berlin-Heidelberg, Springer Verlag, 2001, pp. 1-28.

[24] For an overview of the legal perspective on artificial agents, see G. SARTOR, *Cognitive Automata and the Law: Electronic Contracting and the Intentionality of Software Agents*, in "Artificial Intelligence and Law", Vol. 17, 2009, pp. 253-290; and S. CHOPRA, *Rights for Autonomous Artificial Agents?*, in "Communications of the ACM", Vol. 53, 2010, n. 8, pp. 38-40.

[25] This core content has been extended to also include intelligence and intentionality. The list can be extended even further (see, for example, S. WILLMOTT, *Illegal Agents? Creating Wholly Independent Autonomous Entities in Online Worlds*, 2004, http://www.lsi.upc.edu/dept/techreps/llistat_detallat.php?id=695), but the point, where we are concerned, is not

can interface so as to complement each other's work, in that the concept of autonomy informs the vocabulary of both lawyers and software engineers. Let us see, then, if we can flesh out this idea a little further.

4. Autonomy as an Interface between Law and Computer Science

We will be considering in the rest of this discussion how the concept of autonomy – especially as applied to artificial agents within an MAS – can offer a platform for collaborative action between two areas of practice: law and computer science. To see this, we will first briefly consider the concept of autonomy in philosophy and will then work out a *legal* concept of autonomy so as to set out the conditions satisfying which an agent's action can be deemed autonomous, in effect identifying a threshold of autonomous action that would trigger legal consequences (research line *b* above). I argue that artificial agents have yet to reach this threshold, and I discuss the legal considerations that would have to be taken into account in light of this future development. With that done, I tie this idea of autonomy to the previous discussion on boundary objects by treating autonomy itself as a boundary object and arguing that if law and computer science will look at autonomy in this light, positive results will come out of research on MASs.

4.1. The Concept of Autonomy

One way to broach the subject of autonomy is to consider all the terms that scholars in the humanities have associated with this concept: we find freedom, liberty, independence, choice, decision-making, dignity, integrity, authenticity, the self, individuality, personality, rationality, reflexivity, and strength of will, among others[26]. This gives us a flavor for what autonomy can involve, and since there is much disagreement about the role that each of these terms plays in shaping the concept of autonomy, Dworkin has come to the conclusion that "the only features held constant from one author to

so much to agree on a list of core properties as to see in these properties points of contact enabling different CPs to work on joint projects.

[26] The longer list, with an account of how each of these terms is related to autonomy, can be found in D. Milligan, W. Watts Miller (eds.), *Liberalism, Citizenship and Autonomy*, Aldershot, Avebury Ashgate Publishing Ltd., 1992.

another are that autonomy is a feature of persons and that it is a desirable quality to have"[27].

I take the stronger view of autonomy as a core property of agency, or personhood, because (among other reasons) I am interested in finding a basis on which different communities can work collaboratively in reasoning about agents. For this purpose I turn to Kant, not only because he gave us the modern conception of autonomy such as it applies to us as practical moral agents[28], but also because in so doing he set out a broad framework that can be used to shed light on the different areas of practice where autonomy has a role.

Kant bound together in a tight knot the ideas of freedom, autonomy, and moral personality. He essentially said that to be a moral person (in essence a *person*) is to be autonomous, and that what it means to be autonomous is to govern oneself according to the moral law, a law we receive from our common human reason and impose on ourselves by self-legislation: this is something we *choose* do to, that is, we are free to follow the moral law (freedom of will)[29], but if we choose not to, then we shed our humanity, that is, we can no longer be said to exist as moral agents. Thus moral personality (or personhood) implies autonomy, which in turn implies freedom (free will), but we can only be free if we follow the moral law of common human reason[30]. Two points here bear comment. The first of these is that autonomy is a twofold concept, on the one hand it requires freedom, but then this freedom resolves itself into our ability to follow the moral law (to be our own law-givers), from which it follows that autonomy cannot just be equated with freedom: it is freedom responsibly exercised in accordance

[27] G. DWORKIN, *The Theory and Practice of Autonomy*, Cambridge, Cambridge University Press, 1988, p. 6.

[28] The Kantian account figures prominently, for example in the political philosophy of John Rawls (J. RAWLS, *A Theory of Justice*, Cambridge, Harvard University Press, 1971) and also informs the work of O'Neil (O. O'NEIL, *Autonomy, Coherence and Independence*, in Milligan D., Watts Miller W. (eds.), "Liberalism, Citizenship and Autonomy", cit., pp. 203-225).

[29] "What else can freedom of will be but autonomy – that is, the property which will has being a law to itself?". See I. KANT, *Groundwork of the Metaphysic of Morals*, London-New York, Routledge, 2001 (1st ed. 1785), p. 107.

[30] It is through this common human reason that we can distinguish right from wrong in practical matters: "The commonest intelligence can easily and without hesitation see what, on the principle of autonomy of the will, requires to be done". See I. KANT, *Critique of Pure Reason*, Mineola, Dover Publications, 2004 (1st ed. 1781), p. 38.

with the moral law, thus taking into account the *consequences* of our actions. And the second point is that the moral law, as a principle we find in our common human reason, is something through which we can relate to others according to what is right. The thrust of these two points is that autonomy in Kant can be understood as an inherently *relational* concept: it is not just our choosing a course of action on our own, but our doing so consistently with what *others* are doing.

This shows how the concept of autonomy can easily find its way into the law, for law itself is an inherently relational device, by which I mean that law is the primary technique we have for "getting along" even in contexts where we are perfect strangers to one another. And just as autonomy as a moral concept requires us to follow a moral law that takes others into account, so as a legal concept it can be construed as the basic attribute through which to ascribe responsibility, in that we can only be held responsible for something we did willingly on our own (autonomously), and to be held responsible in the law is to be made to recognize the consequences our actions have on others. But let us see in greater detail what it means to be autonomous in a legal sense, for in this way we can draw up a list of characteristics an artificial agent should satisfy to be counted as autonomous in that sense.

4.2. The Concept of Autonomy in Law

Law was earlier described as "the most pervasive and developed normative system"[31], but the full description reads, "and it is typically concerned with the government of autonomy". This should not strike us as surprising if we consider the idea in light of the foregoing discussion: if autonomy is central to a moral person's agency, and hence to practical reasoning at large, and if practical reasoning informs different areas of human activity – one of which is the law – then it only makes sense that autonomy should also be central to law. This is easily grasped as a general proposition, that law is called on to regulate our interaction as autonomous agents, but how is such autonomy understood as a specifically legal concept?

It should be noted, before we begin, that because we are considering autonomy in law proceeding from a philosophical account of autonomy, that account will inevitably inform the legal concept we wind up having: we will end up with a legal-philosophical rather than a strictly legal concept of au-

[31] R. CONTE, R. FALCONE, *Introduction. Agents and Norms: How to Fill the Gap?*, cit., p. 2.

tonomy considered from a lawyerly legal-positivist perspective. We do this precisely because, in this way – by taking a broader perspective than that of the law itself – we can hope to see solutions that would not otherwise be available from that more circumscribed perspective.

With that said, I begin by noting that autonomy can take on any of several meanings in law depending on the area of law we are dealing with (contract, labor, family, tort, international private and public law, etc.), on who the specific *subject* of law is (a citizen, a worker, a spouse, or an organization or a nation, etc.), and on what interests are being protected (freedom of contract, freedom of speech, occupational safety, etc.). So it is doubtful that there can be such a thing as "*the* legal concept of autonomy", but some commonalities can be found, and indeed Lapidoth[32] has distilled three broad meanings of autonomy in law: autonomy can be understood as (*a*) the right to act according one's own beliefs (positive freedom); (*b*) the ability to act independently of a centralized system (negative freedom, noninterference); or (*c*) the exclusive power of an entity to administer and rule on certain issues (regulatory and adjudicative power).

We can briefly consider the first of these three characterizations of autonomy (the right to act according one's own beliefs), for it seems to act conceptually as a condition for the second and the third one and can thus be construed as more foundational. Three features of this definition stand out. The first of these is that autonomy is defined as a right. This is somewhat startling, as it suggests that we have a right to be autonomous: What is the sense, one might ask, of stipulating a *right* to do what we already have the *power* to do as practical agents? But, as we will see shortly, the bewilderment goes away once we realize that rights are essentially the technical device the law uses to body forth the idea of autonomy: they are the legal implementation of autonomy. The second feature that stands out is how vague the definition is, for it does not specify what it means to act according to one's own beliefs: these could be religious (freedom of conscience) or they could be epistemic beliefs about the world or about what the best course of action is for us to take given what our goals are (autonomy as rationality). But here we can see that the vagueness is grounded in the conceptual difficulty (and inadvisability) of offering a one-size-fits-all definition of autonomy. And this in turn brings out the connection with philosophical accounts of autonomy,

[32] R. LAPIDOTH, *Autonomy: Potential and Limitations*, in "International Journal on Group Rights", 1993, n. 1, p. 277.

with the underlying idea of freedom of choice: consider, for example, Raz, for whom "a person is autonomous only if he has a variety of acceptable options open to him to choose from and his life became as it is through his choice of some of these options"[33]. And the third feature relates to the idea of autonomy as a right: it is widely known that rights in the law are correlative to duties[34], in that one person's right entails another's duty, and this means that autonomy is not just the right to freely make our own choices but also takes into account what those choices entail once acted upon. This goes back to Kant's conception of autonomy as something that binds together freedom and the moral law as a law that takes other agents into account. And the same basic principle is at work in the law as a relational device: the moment we speak of the rights one holds as an autonomous agent, we must thereby consider the consequences that follow from one's use of those rights and the responsibility that comes with that use[35] whence comes the whole cluster of legal concepts that includes liability, tort, damages, redress, restitution, and "making someone whole".

What the three foregoing features point to is that autonomy in law can be constructed as the *legal* implementation of a broad *philosophical* concept of autonomy: a practical agent's autonomy is implemented in the law by ascribing *rights* to that person; the moral power through which we exercise these rights is in law our *legal* capacity; and the same rights exist not in isolation but in the relational context of society, a context which brings into play the *consequences* attendant on our exercise of rights, and which in turn brings in the correlative idea of a duty (e.g., the duty or obligation to pay damages for violating another person's property rights).

[33] The whole line of communitarian philosophy has raised strong objections to the idea of the moral agent as an autonomous individual with a number of options to choose from in the abstract, because every individual is embedded in a social context, and that context acts as a force in shaping our lives and persons. These are serious objections, to be sure, but we need not take them up here because they will not help us clarify the idea of autonomy such as it figures in the working of the law. See J. RAZ, *The Morality of Freedom*, Oxford, Oxford University Press, 1986, p. 204.

[34] See W.N. HOHFELD, *Some Fundamental Legal Conceptions as Applied in Judicial Reasoning*, in "The Yale Law Journal", Vol. 23, 1913, n. 1, pp. 16-59.

[35] Therein lies another similarity with the philosophical account of autonomy. Lucas, for example, argues that "I and I alone am ultimately responsible for the decisions I make and am, in that sense, autonomous" (J.R. LUCAS, *Principles of Politics*, Oxford, Clarendon Press, 1966, p. 101).

Legal capacity is something the law ascribes to natural persons (human beings) or to artificial persons, understood entities created by natural persons and treated by law as if they were actual persons[36]. Examples are corporations and partnerships, but what about artificial agents? Do they have the legal capacity the law ascribes to artificial persons? And does that make them autonomous in the eyes of the law? And what would be the implications of such an ascription? These questions I briefly discuss in the final Section below, where I also underscore that autonomy itself can be construed as a boundary object making it possible for different CPs (lawyers and computer scientists) to interact in working out the legal consequences of developing autonomous artificial agents (research line *b* above, as specified in Section 3.2.) and developing MASs that will simulate the actual working of legal rules (research line *a*).

4.3. The Legal Autonomy of Artificial Agents

In the foregoing discussion we considered two key concepts that, if cast in the right way, can serve as a platform enabling lawyers and computer scientists to work together in thinking about and designing artificial agents and MASs. These concepts are that of an agent and that of autonomy. We worked out the relation between these two concepts by arguing that autonomy can be construed as the core content of any entity which can be described as an agent. It was also argued that in this way we can conceive agents – and by extension *artificial* agents – as boundary objects, because once we have agreed on the idea of an agent as an autonomous entity, we have a criterion on which basis to tell an agent apart from a nonagent, and we can also proceed on this basis to design different types of agents and fit them into a legal scheme.

But it was further pointed out that the inquiry is not complete until we also have a notion of autonomy. We thus considered the idea of autonomy as a philosophical concept providing a basis for understanding what an agent's autonomy might mean in the law. The philosophical concept was analyzed as having the core content of self-government understood as an agent's ability to make its own decision about how to act. And it turned out that there are two sides to this feature of autonomy (the ability to set one's own course of action), for on the one hand it means that autonomous agents enjoy free-

[36] S.v. "Artificial person" in S.M. SHEPPARD (ed.), *The Wolters Kluwer Bouvier Law Dictionary*, compact edition, New York, Wolters Kluwer Law & Business, 2011.

dom of action, but on the other hand this freedom entails responsibilities, in that an autonomous agent is really free to determine its own action, and on a Kantian conception is authentically autonomous, only insofar as it takes into account the way in which *others* stand to be affected by such action. It was finally argued that these features of autonomy have been implemented in what might be described as the *legal* understanding of autonomy (however much this may be a misnomer, considering that the idea of autonomy takes a variety of inflections in the law depending on the area of law involved and on the subject or person to whom autonomy is ascribed): an autonomous person is one who is recognized as having legal capacity, namely, the freedom to do juristic acts understood as acts *in* the law (Section 3.2.), such as voting and entering into contracts; such legal capacity confers rights on the person recognized as having such legal capacity; and with these rights also come duties and responsibilities, because no person in the law is understood as an isolate but as someone living in the interactive context of society.

I should now point out that this makes autonomy itself a boundary object, precisely because we can identify in it a core meaning amenable to different implementations: its core meaning is what enables different CPs to recognize an autonomous agent as such, while its amenability to interpretation allows the same CPs to fashion the concept in different ways. And because none of these inflections can depart so much from the core content of autonomy as to prevent one CP from recognizing another CP's autonomous agent as such, we know that the two CPs can understand and work with each other on common projects.

The projects we are specifically concerned with are those that revolve around artificial agents and MASs, and we can now ask: When can an artificial agent be said to be autonomous in a legal sense? Because depending on how we answer that question we may need to work out a legal scheme through which to regulate the use of artificial agents[37] – an endeavor that would fall under research line *b* above – and it is only through a proper understanding of legal autonomy that software engineers can design MASs capable of simulating acts *in* the and acts *of* the law (research line *a*).

[37] G. SARTOR, *Cognitive Automata and the Law: Electronic Contracting and the Intentionality of Software Agents*, cit.; E.A.R. DAHIYAT, *Intelligent Agents and Liability: Is It a Doctrinal Problem or Merely a Problem of Explanation?*, in "Artificial Intelligence and Law", Vol. 18, 2010, pp. 103-121; F. ANDRADE, P. NOVAIS, J. MACHADO, J. NEVES, *Contracting Agents: Legal Personality and Representation*, in "Artificial Intelligence and Law", Vol. 15, 2007, pp. 357-373.

The first thing to consider as we turn to the question of what attributes an artificial agent will have to exhibit in order to be considered autonomous in a legal sense is what computer scientists themselves think about an artificial agent's autonomy. I would note in this regard that software engineers are not ready to consider artificial agents properly autonomous entities[38]. And I would further note that software engineers have a view of autonomy very much attuned to what was earlier identified as the core idea of autonomy as both a philosophical and a legal concept, namely, the idea of self-government, i.e., choosing one's own course of action. In Sycara, for example, an artificial agent's autonomy is defined as an agent's ability to control its actions and the internal states through which it deliberates on these actions[39]. However, in contrast to the idea we have of the autonomy ascribable to a subject of law having legal capacity, an artificial agent's control over its own internal states is limited to its ability to establish the means of getting to an end without also choosing the end itself[40]. In other words, at the current state of technology, artificial agents are autonomous in determining *how* to do something but not in determining *what* to do.

I believe that in this distinction lies a bright-line test for determining whether an artificial agent can be regarded as autonomous for the purposes of the law: an artificial agent can be so regarded – as an agent with legal capacity – only if designed so that it can make determinations about the *ends* of its own action (rather than only about the means of action). And I further believe that in combination with that criterion we should also consider the *range* of ends an artificial agent can decide to pursue: if that range is a finite, predetermined set, the computational entity in question cannot be deemed autonomous in a legal sense (just as no human person can be deemed autonomous in the law if he or she is acting under duress, a circumstance under which the person would no longer be held accountable for those actions). So, if we combine these two criteria, we would come out with this broad rule: an artificial agent can be deemed autonomous under the law, or an artificial person with legal capacity, if (*a*) it can choose its own ends of

[38] T. SMITHERS, *Autonomy in Robots and Other Agents*, in "Brain and Cognition", 1997, n. 34, pp. 88-106.

[39] K.P. SYCARA, *The Many Faces of Agents*, in "AI Magazine", Vol. 19, 1998, n. 2, pp. 11-12.

[40] D.J. CALVERLEY, *Imagining a Non-biological Machine as a Legal Person*, in "Artificial Intelligence and Society", Vol. 22, 2008, n. 4, pp. 523-537.

action (rather than only the means to such ends) and (*b*) these ends are not a finite set[41].

5. Closing Remarks

If we can agree on a legally relevant idea of autonomy, we will have a basis on which to work in setting out a corresponding set of rights and responsibilities. The two-part test offered in the previous Section can serve as a basis for thinking about such rights and responsibilities (a task that computer scientists have already entrusted to lawyers[42]). And it makes sense to think, as a matter of principle, that the more an artificial entity becomes autonomous – indeed the more it resembles a natural person with the moral power to act *for a reason*, or else to distinguish right from wrong and to act accordingly – the greater will be the cluster of rights and duties ascribable to this entity and to those it interacts with.

But the broader comment I would like to make as we wind down is that a good way to think about these issues is the interdisciplinary approach we can take by viewing our basic concepts as boundary objects: this holds not only for the concepts of agent and autonomy, but also for the related concepts of person, moral capacity, rights, and responsibility, among others. If we can think in those terms, we can take down the barriers that might divide different areas of practice: we can do so by finding a vocabulary that makes sense to the communities in these different areas, and in this way we can foster the sort collaborative enterprise needed to attack problems such as those we have been considering in connection with MASs and artificial agents.

[41] Computer scientists speak in this latter sense of autonomous artificial agents as capable of interacting with their environment (in a reactive as well as in a proactive way) and of learning from experience.

[42] See P.M. Asaro, *What Should We Want from a Robot Ethic?*, in "International Review of Information Ethics", Vol. 12, 2006, n. 6, pp. 9-16.