FIDIS

Future of Identity in the Information Society

Title:	"D17.3: Bridging the accountability gap: rights for new entities in the information society?"
Author:	WP17
Editors:	Bert-Jaap Koops (TILT, Netherlands) Mireille Hildebrandt (VUB, Brussels) David-Olivier Jaquet-Chiffelle (VIP, Switzerland)
Reviewers:	Harald Zwingelberg (ULD, Germany) Ronald Leenes (TILT, Netherlands)
Identifier:	D17.3
Type:	[Report]
Version:	1.0
Date:	28 April 2009
Status:	[Final]
Class:	[Public]
File:	fidis-wp17-del17.3-rights_for_new_entities_def.pdf

Summary

New entities in the information society, such as pseudonyms, avatars, software agents, and robots, create an 'accountability gap' because they operate at increasing distance from their principals. One way of addressing this is to attribute legal rights and/or duties in some contexts to non-humans, thus creating entities that are addressable in law themselves rather than the persons 'behind' them. In this article, we review existing literature on rights for non-humans, with a particular focus on emerging entities in the information society. We discuss three strategies for the law to deal with the challenge of these new entities: interpreting and extending existing law, introducing limited legal personhood with strict liability, and granting full legal personhood. To assess these strategies, we distinguish between different types of persons (abstract, legal, and moral) and different types of agency (automatic, autonomic, and autonomous). We conclude that interpretation and extension of the law seems to work well enough with today's emerging entities, but that sooner or later, attributing limited legal personhood with strict liability is probably a good solution to bridge the accountability gap for autonomic entities; for software agents, this may be sooner rather than later. The technology underlying new entities will, however, have to develop considerably further from facilitating autonomic to facilitating autonomous behavior, before it becomes legally relevant to attribute 'posthuman' rights to new entities.

The FIDIS NoE receives research funding from the Community's Sixth Framework Program

Copyright Notice

PLEASE NOTE: This document may change without notice. Updated versions of this document can be found at the FIDIS NoE website at www.fidis.net.

Members of the FIDIS consortium

1. Goethe University Frankfurt	Germany
2. Joint Research Centre (JRC)	Spain
3. Vrije Universiteit Brussel	Belgium
4. Unabhängiges Landeszentrum für Datenschutz	Germany
5. Institut Europeen D'Administration Des Affaires (INSEAD)	France
6. University of Reading	United Kingdom
7. Katholieke Universiteit Leuven	Belgium
8. Tilburg University	Netherlands
9. Karlstads University	Sweden
10. Technische Universität Berlin	Germany
11. Technische Universität Dresden	Germany
12. Albert-Ludwig-University Freiburg	Germany
13. Masarykova universita v Brne	Czech Republic
14. VaF Bratislava	Slovakia
15. London School of Economics and Political Science	United Kingdom
16. Budapest University of Technology and Economics (ISTRI)	Hungary
17. IBM Research GmbH	Switzerland
18. Institut de recherche criminelle de la Gendarmerie Nationale	France
19. Netherlands Forensic Institute	Netherlands
20. Virtual Identity and Privacy Research Center	Switzerland
21. Europäisches Microsoft Innovations Center GmbH	Germany
22. Institute of Communication and Computer Systems (ICCS)	Greece
23. AXSionics AG	Switzerland
24. SIRRIX AG Security Technologies	Germany

Versions

Version	*Date*	*Description (Editor)*
0.1 – 0.7	19.02.2009 – 09.04.2009	• working versions (all editors)
0.8	10.04.2009	• final version for internal review (BJK)
1.0	28.04.2009	• final version (all editors)

Foreword

FIDIS partners from various disciplines have contributed as authors to this report. The following list names the contributors for the chapters of this report.

Chapter	*Contributor(s)*
Executive Summary	Bert-Jaap Koops (TILT)
Appendix. Text of the Article	Bert-Jaap Koops (TILT) Mireille Hildebrandt (VUB) David-Olivier Jaquet-Chiffelle (VIP)

Table of Contents

Executive Summary

This FIDIS deliverable has been written in the form of a multi-disciplinary academic article. FIDIS authors from different disciplines (regulation of technology, philosophy of law, and mathematics) have teamed up to look at emerging entities in the information society and the 'accountability gap' they create by operating at increasing distances from their principals.

The article builds on the groundwork provided by FIDIS deliverable D17.2, "New (Id)entities and the Law: Perspectives on Legal Personhood for Non-Humans" (Koops and Jaquet-Chiffelle (eds), 2008, available at http://www.fidis.net/resources/deliverables/). D17.2 constituted the first step in FIDIS research into the legal status of new entities in the information society. It followed a bottom-up approach with case studies of pseudonyms, avatars, and software agents, and an analysis of legal aspects from continental legal traditions. D17.3 extends the analysis of D17.2, by a more extensive review of the literature in common-law traditions and including a discussion of robots. It also deepens the analysis by following a more systematic discussion of the two types of personhood: limited (in relation to contracting) and full (in relation to culpability), based on the extended literature review.

By building in this way on D17.2 and re-casting the research findings in the form of an academic article, the authors aim at opening up the results of this multi-disciplinary FIDIS research for a different audience. In particular, the legal academic community is targeted to take note of the research results, since the technical perspective on new entities that is incorporated in this article, is a welcome and important addition to current legal literature. Moreover, the generic nature of the research question and analysis are well-suited for a global audience. To this effect, the article will be submitted to one of the top-5 United States-based law & technology journals.

The text of the article is included as an Appendix to this report. It addresses the following research questions:

> First, given the rise of new types of acting entities in the information society that operate at increasing distance from the persons who employ them, is current law sufficiently equipped to deal with potential conflicts, or would it help to create (limited) legal personhood for some of these new types of acting entities in some contexts?
>
> Second, under which conditions would non-human entities qualify for the attribution of liability based on culpable and wrongful action and under which conditions could such entities claim (post)human rights and liberties?

Based on a review of existing literature on rights for non-humans, with a particular focus on emerging entities in the information society, and distinguishing between different types of persons (abstract, legal, and moral) and different types of agency (automatic, autonomic, and autonomous), the article concludes that interpretation and extension of the law seems to work well enough with today's emerging entities, but that sooner or later, attributing limited legal personhood with strict liability is probably a good solution to bridge the accountability gap for autonomic entities. For software agents, this may be sooner rather than later. The technology underlying new entities will, however, have to develop considerably further from facilitating autonomic to facilitating autonomous behavior, before it becomes legally relevant to attribute 'posthuman' rights to new entities.

Appendix. Text of the Article

Bridging the accountability gap: rights for new entities in the information society?

Bert-Jaap Koops, Mireille Hildebrandt & David-Olivier Jaquet-Chiffelle*

Abstract

New entities in the information society that operate at increasing distance from the physical persons 'behind' them, such as pseudonyms, avatars, software agents, and robots, challenge the law. One way of addressing this challenge is to attribute legal rights and/or duties in some contexts to non-humans, thus creating entities that are addressable in law themselves rather than the persons 'behind' them. In this article, we review existing literature on rights for non-humans, with a particular focus on emerging entities in the information society. We discuss three strategies for the law to deal with the challenge of these new entities: interpreting and extending existing law, introducing limited legal personhood with strict liability, and granting full legal personhood. Full legal personhood implies that entities can be held liable for culpable and wrongful action and can claim (post)human rights like freedom of expression and the right to a fair trial. To assess these strategies, we distinguish between different types of persons (abstract, legal, and moral) and different types of agency (automatic, autonomic, and autonomous). We conclude that interpretation and extension of the law seems to work well enough with today's emerging entities, but that sooner or later, attributing limited legal personhood with strict liability is probably a good solution to bridge the accountability gap for autonomic entities; for software agents, this may be sooner rather than later. The technology underlying new entities will, however, have to develop considerably further from facilitating autonomic to facilitating autonomous behavior, before it becomes legally relevant to attribute 'posthuman' rights to new entities.

Keywords: computer agents, personhood, autonomy, liability, posthuman rights

1.1 Introduction[1]

Technological developments in the information society bring new challenges, both to the applicability and to the enforceability of the law. One major challenge is posed by new entities that operate at increasing distance – in every sense of the term – from the physical persons 'behind' them (the 'principal'), such as pseudonyms, avatars, and software agents. In case of accidents or misbehavior, current laws require the physical or legal person(s) 'behind' the entity to be found so that she can be held to account. This may be problematic if the linkability of the principal and the operating entity is questionable.

In case of a pseudonym, for example of an eBay account, the physical person who uses the pseudonym is legally responsible; however, the law too often becomes useless because it is

* Bert-Jaap Koops, MSc (mathematics) and MA (literature), Groningen University, PhD (law), Tilburg University; Professor of Regulation & Technology, Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, the Netherlands.
Mireille Hildebrandt, LL.M, University of Leyden, PhD (philosophy of law), Erasmus University Rotterdam; Associate Professor of Law, Erasmus School of Law, Rotterdam, the Netherlands, and Senior Researcher at Law Science Technology and Society, Vrije Universiteit Brussel, Belgium.
David-Olivier Jaquet-Chiffelle, MSc (mathematics), PhD (mathematics), University of Neuchâtel; Professor of Mathematics and Cryptology, University of Applied Sciences of Bern, Switzerland; Associate Professor, School of Criminal Sciences, University of Lausanne, Switzerland.

[1] This article was written as part of the EU-funded project FIDIS (Future of Identity in the Information Society), http://www.fidis.net. It builds on previous work in which we co-operated with Harald Zwingelberg, Unabhängiges Landeszentrum für Datenschutz, Kiel, Germany, whom we thank for his help in this research. We also thank Ronald Leenes of Tilburg University for his helpful comments on an earlier draft of this article.

hard to enforce legal rights. Indeed, the link between the physical person and her pseudonym can often not be revealed with a reasonable amount of effort. In case of a software agent, who is the person responsible – its programmer, its seller, or its user? What happens if the software agent adapts itself and learns from its environment, so that it eventually behaves in an intrinsically unpredictable way? Is it then still meaningful to find a physical person or another entity with legal personhood who is accountable for the behavior of this software agent?

One solution to this problem has been much discussed in the literature: could or should we attribute legal personhood to such entities, so that they can be legally addressed themselves? Attributing personhood to non-human 'entities' is not as strange as it might seem at first sight. In most modern legal systems, legal personhood is attributed to associations, funds or even ships, even if this is never full personhood in the sense of an entitlement to claim the entire range of human rights and liberties. However, in principle the law can attribute conditional legal personhood to any well-defined type of entity. Clearly, this does not imply that we can simply give legal personhood to avatars or software agents. The law has a respectable tradition in flexibly incorporating social and technological developments in its system. New conditions created by new paradigms have often successfully been interpreted in terms of the existing legal framework. At the same time, we also see that when this interpretation becomes too difficult or too costly to maintain, the legal system has proven itself dynamic enough to move along with new paradigms: new legal constructions or even new legal entities have been created. For example, legal subjectivity has been granted to non-human entities, such as companies, trust funds, and states.

Now, when an action or a transaction is realized with the help of an intermediate acting entity, and when this action or transaction cannot be linked to the person who is legally responsible today, what are possible solutions to make the law applicable and enforceable? Can current laws comfortably incorporate the new entities, or do we need to use again the dynamism of the legal system to create new legal constructions or even new legal persons?

This issue has been discussed in the literature for almost two decades. Since Lawrence Solum's landmark article on 'Legal Personhood for Artificial Intelligences', technologies have considerably advanced, new entities like avatars have emerged, and the literature has moved along. Recently, an important addition to the literature has been published in German – a dissertation by Andreas Matthias –, which may not yet be familiar to the English-language community. Altogether, there is sufficient reason to conduct a review of the literature in light of the on-going developments in ICT-facilitated acting entities, in order to give a closer look at arguments pro and con legal personhood for some of the non-human acting entities, including a discussion of alternative approaches to solving the 'accountability gap'. The research questions we aim to answer in this article are the following.

> First, given the rise of new types of acting entities in the information society that operate at increasing distance from the persons who employ them, is current law sufficiently equipped to deal with potential conflicts, or would it help to create (limited) legal personhood for some of these new types of acting entities in some contexts?
>
> Second, under which conditions would non-human entities qualify for the attribution of liability based on culpable and wrongful action and under which conditions could such entities claim (post)human rights and liberties?

Given the generic nature of these questions, we focus on law in general rather than on specific legal systems, and we do not aim at providing a definitive answer to this question. Rather, we give various perspectives from common-law and continental traditions that are relevant for answering it, in order to come to a tentative conclusion on which future research can build. We start with an introduction to the challenge of various entities operating at increasing distance from their users (§2) and a clarification of the concepts of persons, agents, and autonomy (§3). Next, we provide an extensive review of literature on the topic of rights for

non-humans, from the landmark analysis of Solum to recent literature from Germany (§4). After distinguishing between various types of personhood and agency that emerge from this review (§5), we answer the research question by outlining a three-stage strategy for the short, middle, and long term (§6). We wrap up with a conclusion (§7).

1.2 Facing the challenge: emerging entities in the information society

1.2.1 Pseudonyms

The term "pseudonym" comes from the Greek word *pseudonumon* which means *false name*. Traditionally, a pseudonym refers to a fictitious name taken by an author, a pen name. *Voltaire* and *Molière* are pseudonyms of famous French writers. Nowadays, pseudonyms are often used by artists, especially in show business, to mask their official identity. In this case, a pseudonym can be seen as a self-chosen name becoming an identity in the artist context. In some situations, the pseudonym is used to conceal the true identity of the person, i.e., it acts as a privacy-enhancing tool. Journalists sometimes use such pseudonyms. Pseudonyms also intervene as user IDs in the information society. On the Internet, many people use a pseudonym (or multiple pseudonyms) hoping to stay anonymous. These examples illustrate that a pseudonym, as a (false) name, functions as an 'identity' in common speech and that it can act as a privacy-enhancing tool. Although pseudonyms have a more instrumental, passive nature than the software agents and robots discussed below, they do have a certain independence, because they shield the persons behind them. In a functional sense, the pseudonyms 'do business' on behalf of the person they shield. From this perspective, they constitute an entity in their own right, and it is this abstract entity that makes them a category to consider in our discussion of new entities in the information society.[2] For practical reasons, in this article we will use the term 'pseudonym' as a synonym for the abstract entity that is represented by the pseudonym.

When a pseudonym is functioning as a mask between a human person and the outside world, the pseudonym can acquire a 'personality' of its own and operate at some distance from the person it shields. The user of the pseudonym is connected to the pseudonym (more precisely, to the entity constituted by the pseudonym), but a considerable distance can exist between the user and the pseudonym, precisely because the pseudonym can function in practice as a 'stand-in' for the user with a personality of its own. This is particularly the case when the pseudonym is a mask shared by more than one person, so that it functions relatively independently from the specific human being(s) behind it. A clear example of this is the pseudonyms used on eBay. One of eBay's specificities is to allow users to interact between each other using pseudonyms. During the registration phase, after having given his personal identifying information, the user is asked to create his eBay user ID – a self-chosen pseudonym – that will act later as an anonymizing tool towards the other users.

It is interesting to observe that mechanisms can be developed to deal efficiently and securely directly with the pseudonym itself (i.e., the abstract entity constituted by this

[2] This is in line with the approach proposed by a model based on virtual persons, developed in Jaquet-Chiffelle, D-O. (ed.), *D2.13: Virtual Persons and Identities*, FIDIS deliverable, March 2008, available at http://www.fidis.net (last accessed 28 April 2009). In this model, a pseudonym can be seen as the identity – and identifier – of a virtual person, which is a special type of abstract entity. According to this model, the pseudonym is the tautological identity of its corresponding abstract entity: by definition, it identifies this abstract entity. In other words, the pseudonym represents an abstract entity that is identified by the pseudonym. For example, George Eliot – a pseudonym used by Victorian author Mary Anne Evans – identifies an abstract entity: 'the abstract entity called George Eliot'. This abstract entity does not exist as a person of flesh and bone, but is a virtual person known to many as, for example, the author of *Middlemarch*.

pseudonym) rather than the individual using this pseudonym. Payment procedures and reputation on eBay are good examples. Different payment methods are available on eBay.[3] Every seller can choose the methods of payments that he offers. Most sellers offer PayPal which is the preferred payment method for most eBay users. The PayPal method presents several advantages in comparison to the credit card method. The credit card used by PayPal (if a credit card is used[4]) is not divulgated to the seller. Therefore, the buyer only needs to *trust* PayPal (not the seller himself anymore) not to misuse his credit card information. The reputation of PayPal is currently more reliable than the reputation of any specific seller. Moreover, PayPal offers "PayPal Buyer Protection"[5] that "helps you if you don't receive your item or the item is significantly different from its description in the seller's listing."

Reputation is a key component when building trust. PayPal, for example, is more trusted than any escrow service in particular because it has a strong implicit positive reputation, just by being the preferred payment method for most eBay buyers and sellers. The eBay platform provides a reputation system that allows building trust between eBay users who do not know each other, who have never interacted together and who are hidden behind pseudonyms. To each eBay user ID is attached a so-called "feedback profile". The feedback profile of an eBay user ID measures the concordance between the actual behavior of this eBay user ID during his previous transactions and the expected behavior of this eBay user ID, according to other users who have already taken part into these transactions. The eBay reputation system is fed by users themselves. It collects experiences of previous eBay transaction partners.

One thinkable advantage of addressing a pseudonym as the identity of an intermediate entity rather than identity-related information of the person behind it only, is the possible privacy-enhancing effect. By hiding the link between the intermediate entity and the physical person, anonymity for the physical person may be achieved if this is acceptable for the parties involved.

For acceptance in commercial and legal practice, the ability to "de-anonymize" is currently an important attribute of pseudonyms. A pseudonym is "de-anonymizable" when the information that provides the link to the physical person can be disclosed upon request. This leads to the difference whether or not the identity of the holder of the pseudonym can be disclosed at least in a defined set of situations, for example when a contractual party does not comply with its duties. Such disclosure as well as the control over the requirements of a disclosure may be handled by a trusted third party, a linkability broker, which needs to be in possession of the information to match the pseudonym with the name of the holder, i.e., the physical person behind the pseudonym.

In trade and private law, trust is a crucial factor influencing the potential use of pseudonyms. Pseudonymous transactions are likely to be accepted in cases of an immediate performance, thus avoiding the need for credit for the holder of the pseudonym; in cases where payment and performance are not simultaneous, the seller needs to trust that payment will follow and the buyer that the product or service will be delivered. Some technical and organizational solutions are available for enhancing trust in these cases, for example PayPal being used in eBay transactions. Before pseudonymous transactions can really flourish, more such trust-enhancing mechanisms will need to be developed and put to practice.[6]

---

[3] http://pages.ebay.com/help/pay/methods.html (last accessed 1 April 2009).

[4] The buyer can choose not to give his credit card number to PayPal and make a fund transfer from his bank account.

[5] http://pages.ebay.com/help/buy/paypal-buyer-protection.html (last accessed 1 April 2009).

[6] See, e.g., J.E.J. Prins et al., *Trust in Electronic Commerce: The Role of Trust from a Legal, Organizational and Technical Perspective*, Kluwer Law International, 2002; D.-O. Jaquet-Chiffelle and H. Buitelaar (eds.) (2009), *Trust and Identification in the Light of Virtual Persons*, FIDIS deliverable D17.4, available at http://www.fidis.net/resources/deliverables/ (last accessed 28 April 2009).

1.2.2 Avatars

Avatars are entities featuring in computer games and other online environments like Second Life, for example. Such digital avatars represent the player in the game world of Multi User Dungeons (MUDs), Multi User Virtual Environments (MUVEs), Massively Multiplayer Online Role Playing Games (MMORPG) and other computer games (further referred to as "virtual games"). The term avatar does not only refer to three-dimensional representations in virtual games but also to icons representing a specific user in an online forum or any other graphical representation of a computer user.[7] For our purposes, an avatar is a virtual person representing one or more players in the physical world or even a computer program, as used by game publishers to control non-player characters (NPCs).

Engaging in a virtual game usually starts off with the creation of a personalized avatar by adjusting the appearance of the graphical representation on the screen by choosing skin, facial features and clothes. In many games, particularly role playing games, further attributes such as strength, dexterity and abilities such as swimming, climbing or pickpocketing can be assigned to further personalize the avatar. These attributes may then be improved during the course of the game allowing the avatar to act more efficiently. In fact, in many role playing games advancement and development of the avatar is a central aspect of the game play. Guiding an avatar in its advancement over a long time and individualizing the avatar with one's own preferences or getting absorbed by the interaction with other avatars forges a tight relation between the player and his avatar.[8]

As having an advanced avatar makes the game play more enjoyable, a demand for both good items and well-developed avatars as a whole exists, creating a market for such virtual goods. Depending on the game publishers' terms of service such trade may be allowed, even intended, limited to in-game trade, or forbidden. Increasingly, publishers allow and encourage the transfer of avatars between players. The increased demand and market value of virtual items gave rise to legal discussions and has even led to first legal actions brought to national courts.[9]

Contrary to certain types of pseudonyms (*supra*) and software agents (*infra*), avatars are not usually involved in commercial relationships but rather in leisure contexts. As such, their legal status is relevant in light of the tight emotional bond which physical persons can establish with their avatar.[10] This raises the question whether, for example, defamation of an avatar can occur and if so, whether it has legal consequences. Are only physical persons a qualified target or is it possible to defame a virtual entity such as an avatar as well? Based on its prior actions, an avatar may have a reputation of some kind. Part of this reputation may be represented by a ranking or reputation system within the virtual world. The programs and scripts in control of other avatars could refer to this kind of reputation of the avatar to calculate their response towards the avatar. Such reputation may even become a factor affecting the economic value of the avatar in the physical world. Damaging this reputation could cause a monetary loss for the player in the physical world, for instance in case he wants to sell his avatar, and this may constitute a tort and could lead to granting a claim for damages. But in contrast to reputation, an avatar is not capable of having honor, dignity, or

---

7 http://en.wikipedia.org/wiki/Avatar_%28computing%29.

8 Yee, N., 'The Psychology of Massively Multi-User Online Role-Playing Games: Motivations, Emotional Investment, Relationships and Problematic Usage', in: Schroeder, R., Axelsson, A., (eds.) Avatars at Work and Play Collaboration and Interaction in Shared Virtual Environments, Springer Netherlands, 2006, pp. 187-207, available online at http://vhil.stanford.edu/pubs/2006/yee-psychology-mmorpg.pdf.

9 A recent case concerned the sale of a piece of virtual land. See *Bragg* v. *Linden Research, Inc. et al.*, http://docs.justia.com/cases/federal/district-courts/pennsylvania/paedce/2:2006cv04925/217858/26/.

10 See Jaquet-Chiffelle, D-O. (ed.), *D2.13: Virtual Persons and Identities*, FIDIS deliverable, March 2008, available at http://www.fidis.net, p. 12.

self-esteem. Consequently this raises the fundamental question as to what exactly is the object of the protection offered by the regulations on defamation in the different jurisdictions.

1.2.3 Software agents

In the information society, more and more tasks are facilitated, and indeed increasingly performed, by software. As the software program becomes more autonomous, we can speak of software agents. These are sometimes also referred to as electronic agents, intelligent agents or softbots (software robots), although some scholars use these terms for specific kinds of software agents.

To illuminate the concept of software agents, it is useful first to look at the concept of an agent. Generally speaking the term 'agent' refers to:[11]

1. an entity capable of action;[12]
2. someone (or something) who acts on behalf of another person.[13]

In the first, most general, sense, the class of agents can be divided into biological agents (such as human beings or viruses) and non-biological agents, which include both hardware agents or robots and software agents. All of these are capable, to a larger or smaller extent, of action. If the action is performed on behalf of another entity, the second, more restricted, sense is activated: the agent then functions as a representative of another entity. In this section, we focus on software agents, in both senses of the term.

If we restrict the notion of action to intentional or autonomous action, not all software qualifies as an agent in the sense of an entity capable of action. 'Software agents are programs that react autonomously to changes in their environment and solve their tasks without any intervention of the user.'[14] Because of this characteristic, software agents are sometimes also called autonomous agents.[15] Note that in this definition, intention is not required and autonomy is understood in a very general manner that includes 'actions' of 'agents' that are not aware of their own actions and for this reason cannot be held morally responsible for them. Below, in section 1.3, we shall further explore the nexus of agents, autonomy and personhood.

Various kinds of software agents exist. A distinction can be made between stationary agents and mobile agents. Stationary agents move only in their original environment (e.g., their owner's computer), whereas mobile agents 'move around (migrate) independently in heterogeneous computer networks'.[16] Agents can also be classified according to their function. There are basically four types of software agents:[17]

1. user agents (personal assistants);
2. buyer agents (shopbots);

---

[11] http://en.wikipedia.org/wiki/Agent. See *infra*, s. 1.3 for further elaboration.

[12] *Cf.* Webster's Online Dictionary: 'An active and efficient cause' or a 'substance that exerts some force or effect'. More interesting for our purpose is Latour's definition of an actor: '*any thing* that [modifies] a state of affairs by making a difference' (emphasis in the original). B. Latour (2005), *Reassembling the Social. An Introduction to Actor-Network-Theory*, Oxford: Oxford University Press at 71. See also *infra*, s. 1.3.

[13] *Cf.* Webster's Online Dictionary: a 'representative who acts on behalf of other persons or organizations'.

[14] S. Wettig and E. Zehendner (2004), 'A legal analysis of human and electronic agents', 12 *Artificial Intelligence and Law* 111 at 112.

[15] 'An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.' S. Franklin and A. Graesser (1996), 'Is it an Agent, or just a Program? A Taxonomy for Autonomous Agents', in: *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*, Berlin: Springer-Verlag 1996. Note that we use the term autonomous in a more restricted sense, see *infra*, s. 1.3.2.

[16] Wettig and Zehendner, *supra* note 14 at 112.

[17] http://en.wikipedia.org/wiki/Software_agent.

3. monitoring or surveillance agents;
4. data mining agents.

User agents are typically stationary and restricted to personal use; as a result, they raise less questions with regards to duties and obligations than the other types, which are usually mobile and therefore more distant from their owners. In terms of 'distance' from their principal, it is also useful to distinguish three types of agents, depending on the degree of autonomy with which they operate.

- **A slave**: a slave has no autonomy at all. For any decision that affects the possessions, legal rights and obligations of its 'master', it has to consult him.
- **A representative:** it may take its own decisions within a well-defined domain and within strict limits.
- **A salesman**: this agent may take its own decisions and is not restricted in the way in which it intends to take care of its user's interest. It is bound to serve the interests its user wants to be taken care of. It may for instance manage a stock portfolio belonging to its user.

It is because of their relative autonomy that these software agents are relevant to study from the perspective of the law: they are (normally) related to physical persons, but at a distance, and hence time may come when their actions can no longer be seen as the actions of the human beings 'behind' them. In as far as these actions have legal or other consequence, this raises the issue of whether and to what extent rights and obligations should be attributed to software agents themselves. This is a highly relevant question in an information society in which these agents become increasingly autonomous. Indeed, if we are to believe Willmott, 'it may now be possible (…) to construct wholly independent autonomous electronic entities able to act for themselves in the real world: sustaining themselves financially, possessing their own identity and surviving unaided for periods of up to several years'.[18]

1.2.4 Robots

Long before the notion of software agents emerged, the idea of autonomic machines – robots – was already prevalent, first in fiction and, with slowly increasing sophistication, in reality.[19] Karl Čapek introduced the term 'robot' in his 1921 play *R.U.R.* (*Rossum's Universal Robots*), for servant machines looking like humans. Most robots in real life are industrial robots, used for example in car and electronics factories, but service robots, like vacuum-cleaning or lawn-mowing machines, are also being offered on the market. These robots are more than just machines, in that they usually have some sensors for scanning and adapting movements to their environment. They operate without direct human intervention, and appear to have some form of agency. Increasingly, these machines are becoming more autonomic, performing more complex tasks based on programmed algorithms while processing multiple sensory input from their environment.

Another type of robots emerging are pet robots. The tamagotchi, developed in the 1990s, was a primitive and briefly highly popular gadget posing as a pet. Apart from such digital pets, animal-look-alike pets are also being produced. The best-known pet robot is probably Sony's Aibo, a robot-dog introduced in 1999. Paro, a robot seal, for example, is popular in Japan as a pet companion, and he is proposed for therapeutic purposes in hospitals.[20]

---

[18] S. Willmott (2004), 'Illegal Agents? Creating Wholly Independent Autonomous Entities in Online Worlds', Barcelona: LSI, 3 Feb 2004, 8 p., http://www.lsi.upc.edu/dept/techreps/llistat_detallat.php?id=695.

[19] For a good overview, see http://en.wikipedia.org/wiki/Robot.

[20] See, e.g., 'Robot baby seals to replace cats and dogs as pets in hospitals, nursing homes', *Canadian Press* 11 January 2009.

Other types of robots are being developed that begin to look more and more like humans. One strand of research is developing realistic looking robots that mirror human looks.[21] Another strand looks at distinguishing features that can make a robot to be perceived as human, in particular facial expressions like smiling or raising eyebrows.[22] If the 'humanoid' robot were to be equipped with artificial intelligence – and thus acquire more autonomy through emergent behavior –, we are slowly getting closer to the futuristic vision of an android.[23]

Because of the huge potential benefits of automating tasks, the first type of robots (industrial and service) will almost certainly continue to be developed with growing sophistication and an increasing level of autonomic functioning. The development of animal and human-looking robots will also move forward, perhaps with lower levels of autonomic activity than the functional robots because they have a largely social or entertainment function.

1.2.5 Increasing distance

To summarize how new acting entities, as described above, operate at increasing distance, we propose two open questions that illustrate this new paradigm and how this creates problems for legal accountability.

The widespread use of persistent pseudonyms on the Internet, for example of an eBay e-tailer or consumer, raises questions about the link between a transaction and the physical person with whom the transaction is made, since this person is often invisible for most observers. How do we deal with this new reality, when if something goes wrong, no physical person can be linked with a reasonable amount of effort to the transaction? Even if substantive law provides a clear answer who is responsible and who should bear the consequences – which will often but not always be the case with the entities discussed –, can rights be effectively enforced in practice? New forms of unlawful activities take advantage of the gray zones, where the law is theoretically applicable but becomes very hard to enforce in a globalized cyberworld with entities acting at increasing distance.

In order to assess responsibility, the reason why an action took place sometimes has to be determined. Was it done, for example, with *mens rea* (a 'guilty mind')? What happens when a non-human entity acts on behalf of a human being, i.e., when the human being is only indirectly acting at a considerable distance, so that the reason behind the action becomes fuzzy in the translation process from user command to agent act, or becomes difficult to determine when the user cannot be traced with reasonable effort? Can non-human entities, like a software agent, be considered to have their own will and take independent decisions?

1.3 Persons, agents, and autonomy

The entities discussed above do not *prima facie* count as 'persons' – a term traditionally associated with human beings. Yet the artifact of the non-human legal person shows that also non-human entities can count as 'persons' in law. Before moving into a discussion of new

---

[21] See, e.g., the work of Hiroshi Ishiguro, http://www.irc.atr.jp/~ishiguro/.

[22] See, for example, MIT's KISMET, http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html.

[23] The best-known example of an android is Data from StarTrek. For robots to become more 'intelligent' and 'social', context-sensitivity seems pertinent, which could imply, for example, some form of distributed intelligence that emerges less from a single robot than from its interconnectivity (online connections with data bases that allow for data mining the data gathered by both the robot and other sensors in its environment). This could make it hard to identify the robot as a physical entity, as its emergent behavior depends on the entire network of interconnected sensors, and online data bases. See Karnow on similar problems with polymorphous mobile computer agents in s. 1.4.3.1.

types of legal persons, it seems useful to briefly clarify what is meant with a person, a concept that relates to the concepts of agent and agency. These concepts, in turn, relate to varying degrees of automation and autonomy, which are also important to distinguish conceptually.

1.3.1 Personhood and agency

To provide some conceptual coherence, we may start with Bruno Latour's salient depiction of what he calls 'actants':[24] '*any thing* that [modifies] a state of affairs by making a difference'.[25] Any thing *can* thus be an actant in this very broad sense, depending on whether it does or does not make a difference. Paraphrasing Putnam's famous translation of Peirce's pragmatist stance on doubt (one cannot doubt everything, but we should be willing to doubt anything), we could say that 'it makes no sense to qualify everything as an actant, but we should be willing to qualify anything qualifying as an actant'. When discussing legal personhood for non-human actants the point should be to investigate at what point it makes sense to attribute legal consequence of the actants' 'actions' to the actants themselves, instead of to the human actant(s) behind them. In the case of corporations, funds and associations this question has been answered in detail in the positive law of most modern legal systems. To answer this question with regard to pseudonyms, avatars, software agents or robots, we need to establish the conditions under which such attribution solves problems without creating even greater ones. Depending on how novel legal persons are introduced, they could in fact destabilize familiar notions of responsibility that form the moral core of the law, reinforcing undesirable affordances[26] of an increasingly independent technological infrastructure. Instead of reinforcing independent actions of novel technologies over which we have little control, one could also seek protection in the law against what some would qualify as a marginalization of human agency. In this article we shall not assume that technologies are either good or bad, rejecting both techno-optimism and techno-pessimism. Nevertheless, we believe that the emerging proliferation of electronic agents and other quasi-autonomous agents challenges the present legal framework, requiring an in-depth study of the conditions for legal personhood in an information society. This will require the development of a generic vocabulary that takes into account the specificities of both the domain of computer science and of law.

In computer science an agent has been defined as:

> A program that performs some information gathering or processing task in the background. Typically, an agent is given a very small and well-defined task.[27]

Importantly:

> In computer science, there is a school of thought that believes that the human mind essentially consists of thousands or millions of agents all working in parallel. To produce real artificial intelligence, this school holds, we should build computer systems that also contain many agents and systems for arbitrating among the agents' competing results.[28]

Interestingly, in law, an agent is often defined as:

> A person authorized to act for and under the direction of another person when dealing with third parties. The person who appoints an agent is called the principal. An agent can enter into binding agreements on the principal's behalf and may even create liability for the principal if the agent causes harm while carrying out his or her duties.[29]

---

[24] Latour, *supra* note 12 at 71.
[25] *Ibid.*, emphasis in the original. See also *Ibid.* at 52-54.
[26] An affordance can be described as 'what is afforded by a particular technological device or infrastructure'. The term was coined by Gibson in J. Gibson (1986), *The Ecological Approach to Visual Perception*, New Jersey: Lawrence Erlbaum Associates. See also http://en.wikipedia.org/wiki/Affordance.
[27] See http://www.webopedia.com/TERM/A/agent.html.
[28] See http://www.webopedia.com/TERM/A/agent.html.
[29] See http://www.nolo.com/definition.cfm/term/688A86E9-01FC-4A61-BF31C4A315325DAC.

What we basically see here, is that both in computer science and in law the concept of an agent refers to an entity that is at work for somebody (or something) else. In both cases we have a principal that determines the objective, task, scope, means, restrictions, etc. of the agent that he employs. We will therefore refer to electronic pseudonyms, avatars, software agents and robots that act or interact with others on behalf of their users/owners as 'computer agents'. In the present legal framework, a computer agent cannot play the role of a legal agent, because to be a legal agent, the computer agent must have legal personhood; so far, only natural persons, specific types of companies, associations, a trust fund, and public bodies have been attributed legal personhood. If a computer agent would become a legal agent, it could conclude contracts in the name of the principal. In case the agent lacks proper authority of the principal or the principal is non-existent, the contracting partner would be able to sue the agent for breach of contract. One could imagine a restricted kind of legal personhood for computer agents, enabling both the user/owner and those interacting with these agents more leeway in the handling of their affairs. In as far as the interactions initiated by computer agents cause serious harm, we may want to sustain the possibility to attribute legal responsibility for wrongfulness and *mens rea* to actants capable of reflection and intentional action. The notion of calling a person to account for her actions seems to fall flat on its face if applied to contemporary computer agents, and this is one of the issues we will investigate in the following section.

In ethics and philosophy 'agency' is a term reserved for the capability of a person to have intentions and to make conscious deliberate choices on the basis of a moral and/or pragmatic judgment about what is at stake.[30] Even if it makes sense to argue that non-human entities 'act' and make a difference, this is not meant to suggest that they act on the basis of conscious reflection. In as far as legal liability builds on this notion of agency, we need to inquire further into the nature of computer agents and decide whether and when they qualify for such agency.

Personhood is not equivalent with agency, though it is obviously related. Again, in different domains it has different meanings. In computer games a *persona* is equivalent to an avatar, while in legal theory a *persona* is often described as the mask of legal personhood that allows an entity to act in law, while protecting the physical person or other entity behind the mask from being equated with its legal role.[31] The similarity between a persona/avatar and a legal person can be found in the fact that both refer to a role instead of the entirety of a physical entity. This, however, does not imply that the usage of the term is similar in other ways. An avatar-*persona* is created in order to play in a virtual game or roam about in a virtual world; contrary to a legal persona it is not created to provide legal rights and obligations that allow for legal certainty and legal equality. Legal personhood attributes a specific type of personhood to an entity. This notion of legal personhood is related to agency because it enables an entity to act (in law), meaning that the law attributes legal consequences to the actions of the entity. So, if agency refers to an entity's capacity to 'act', to make a difference, legal personhood refers to the fact that this difference generates legal consequences. However,

[30] See http://plato.stanford.edu/search/searcher.py?query=agency for an overview of the intricacies of the concept of agency in law and moral philosophy.

[31] This double function of legal personhood has been further developed in the relational conception of law that sees law in a constitutional democracy as always both instrumental for societal order and protective of individual freedom and the freedom to resist dominant frames of interpretation. See Foqué, R. and A. C. 't Hart (1990), *Instrumentaliteit en rechtsbescherming*, Arnhem/ Antwerpen: Gouda Quint Kluwer Rechtswetenschappen; Gutwirth, S. (1993), *Waarheidsaanspraken in recht en rechtswetenschap*, Brussel: VUB-press and MAKLU; Hildebrandt, M. (2006), 'Trial and "Fair Trial": From Peer to Subject to Citizen', in: *The Trial on Trial. Judgment and Calling to Account*, A. Duff, L. Farmer, S. Marshall and V. Tadros (eds), Oxford and Portland, Oregon, Hart, 2: 15-37.

in as far as the law attributes liability for wrongful actions committed with *mens rea*, another notion of personhood is at stake. This notion of personhood relates to an ethical and philosophical notion of agency that refers to the capacity to act in the sense of intentional meaningful action. Such personhood suggests a sense of self, a capability of standing trial, i.e. of being called to account for one's actions.

One of the pertinent issues that is at stake in this article is the question when legal personhood should be attributed to entities devoid of 'agency' in the ethical and philosophical sense of being capable of intentional action. The problem with the attribution of legal personhood to such entities (animals, ships, trust funds, organizations) is threefold. First, in a court of law they will always have to be represented by entities with agency (at this point in time that means they need representation by human beings). Second, it is difficult if not impossible to establish liability for intentional wrong-doing or criminal guilt in the case of an entity without such agency, which usually means that in those cases the liability of other legal subjects (with such agency) needs to be established.[32] Third, the attribution of legal personhood could entail an appeal to human rights on behalf of the novel legal person, which would be problematic if this entity is not capable of self-reflection.

1.3.2 Automatic, autonomic, and autonomous agents

At this point, it is important to make some conceptual distinctions between different levels of automation and autonomy. For this purpose we will distinguish between automatic, autonomic and autonomous agents. *Automatic agents* refer to the traditional association of automation with mechanical, non-creative applications that perform one or more actions automatically, i.e. in a predefined manner. In software programs automation builds on the application of an algorithm that defines the behavior of the program. *Autonomic agents* refer to some of the entities discussed above[33] that have the capacity to initiate a change in their own program in order to better achieve a certain goal. The program's actions are not entirely predictable, not defined in a closed manner and can thus be said to be *under*determined. Autonomic behavior does not entail consciousness or self-consciousness. *Autonomous agents* refer to those having the capacity to determine their own objectives as well as the rules and principles that guide their interactions. *Auto* (Greek for self) and *Nomos* (Greek for law) refers to an entity capable of living up to its own law. An autonomous agent in this sense is an agent in the traditional ethical and philosophical sense of the term, requiring both consciousness and self-consciousness, i.e., the capacity to reflect upon one's actions and to engage in intentional action. Self-consciousness as the precondition for autonomous action is typical of human agency. So far, machines have not developed consciousness,[34] let alone self-consciousness,[35] while animals with a central nervous system do have consciousness but lack the type of self-

[32] For an interesting brainstorm on the legal personhood of personae without agency see: http://identityblog.burtongroup.com/bgidps/2006/11/the_limited_lia.html and http://thestateofme.wordpress.com/2008/01/09/persona/.

[33] *Supra*, s. 1.2.

[34] In the cognitive sciences, a lively debate discusses whether machine consciousness is possible, how we could design it, and how we could detect it. Leading AI philosophers like Daniel Dennett, who endorse a computationalist understanding of the human mind, see no inherent obstructions to assume machine consciousness is possible, whereas other philosophers within the field of cognitive sciences, like Searle, take a more prudent approach. For an overview, see Holland, O., ed. (2004), *Machine Consciousness*, Exeter, UK: Imprint Academic.

[35] Note that some philosophers, notably Dennett, argue that self-consciousness – if not consciousness itself – is an illusion. This raises the question of the relationship between first person experience and scientific inquiry. See Woodruff Smith, D. and A.L. Thomasson, Eds. (2005), *Phenomenology and Philosophy of Mind*, Oxford University Press.

consciousness that enables reflection and deliberation.[36] Such self-consciousness depends, among other things, on the externalization and constitution of thoughts by means of symbolic language. Evidently, although at present self-conscious machines do not exist, we cannot be sure whether – and if so, when – machines will develop the type of self-consciousness that allows for autonomous action.

1.4 Reviewing the literature: attributing legal personhood?

Legal personhood indicates the capability to be a subject of rights and duties. Within the present legal framework all humans have been attributed legal personhood. It is granted by Art. 6 of the Universal Declaration of Human Rights of 1948 and Art. 16 of the International Covenant on Civil and Political Rights of 1966 to all (living)[37] human beings.[38] The drafters of the European Convention on Human Rights (ECHR)[39] did not provide a similar clause, as they held it to be too trivial and self-evident to include a provision on legal personhood of humans.

All Western legal systems grant legal personhood not only to humans but also to what is called legal persons. Those are legal entities allowing several associations of persons, a trust or even a ship to act in law as if they were a single person (for example registered associations and companies). To protect trade from incapable or fraudulently acting entities, usually high requirements with regard to publicity of the incorporation act apply encompassing mandatory requirements as to formal registration procedures in public registers and mostly some kind of minimum capitalization. This kind of legal personhood is, in opposition to the legal personhood of humans, not attributed by means of international treaties, but rather determined by national law. Within legal philosophy, moral personhood is often seen as precondition for legal personhood, building on French's seminal article on the moral personhood of corporations.[40] French discusses why conglomerates, like a corporation, should be treated as full moral persons, whereas aggregates, such as a lynch mob, do not qualify as such.[41] French distinguishes between metaphysical, moral and legal persons pointing out that for many authors legal personhood depends on metaphysical and/or moral personhood. Obviously current positive law does not concord with this, since no serious argument can be made that a ship or a trust fund is either a metaphysical or a moral person. We therefore refer to the idea – mentioned above – that legal personhood is attributed to enable an entity to act in law (i.e. to create legal consequences) and to be held accountable for its actions, while also protecting the entity itself from being equated with the role it plays. Currently, all entities besides humans and those legal persons recognized by law are considered to be legal objects. This applies,

---

[36] Whether there is continuity or discontinuity between humans and other animals in this respect, see Cheung, T. (2006), 'The language monopoly: Plessner on apes, humans and expressions', 26 *Language & Communication* 316, and De Waal, F. (1997), *Good natured. The origins of rights and wrong in humans and other animals*, Harvard University Press.

[37] For a comparison of the fuzzy borderline at the very beginning of life in German, English, American, French, and Spanish law, see Mahr, J. T., *Der Beginn der Rechtsfähigkeit und die zivilrechtliche Stellung ungeborenen Lebens: Eine rechts vergleichende Betrachtung*, Lang, Frankfurt, 2006.

[38] International Covenant on Civil and Political Rights, U.N. Doc. A/6316 (1966), 999 U.N.T.S. 171 available online at: http://www.unhchr.ch/html/menu3/b/a_ccpr.htm.

[39] The European Convention on Human Rights of 1950, 213 U.N.T.S. 222, available at http://conventions. coe.int/Treaty/en/Treaties/Html/005.htm.

[40] French, P. A. (1979), 'The Corporation as a Moral Person', *16 American Philosophical Quarterly* 207. Qualifying an entity as a moral person does not depend on positive law, whereas qualifying as a legal person obviously does.

[41] *Cf.* Pfeiffer, R. S. (1990), 'The central distinction in the theory of corporate moral personhood', 9 *Journal of Business Ethics* 473.

despite an ongoing movement by animal law activists,[42] also to animals, which are treated as things in private law, being the objects of the rights of their owners.

With the advent of computer agents that operate at increasing distance from their owners, resulting in an 'accountability gap',[43] various authors have discussed the question whether new entities could or should also be attributed legal personhood. If companies and associations can be legal persons, why not software agents as well? In this section, we provide a review of what we consider the most seminal literature that has been published on this question in the past two decades.[44]

1.4.1 Setting the stage: Solum (1992)

In a ground-breaking article in the early 1990s, Lawrence Solum discussed 'Legal Personhood for Artificial Intelligences'.[45] Though technological devices and infrastructures have developed exponentially since he wrote his article, his comprehensive approach is equally relevant today, and we will follow his arguments to see how they can inform us of the conditions under which and the extent to which it makes sense to attribute legal personhood to automatic or autonomic devices or even to non-human autonomous persons.

Solum does not speak of computer agents but of artificial intelligences (AIs). Apart from the pseudonyms, the computer agents described above would qualify as an AI in Solum's terms. At the time he wrote his article, AI was as controversial as it is now.[46] In speaking of AI, we do not take sides in the debate of whether 'non-human intelligence' is a *contradictio in terminis*. We will follow Solum's pragmatic approach, avoiding questions such as 'whether artificial intelligence is possible'. Instead of entering metaphysical debates about the nature of intelligence, his essay 'explores those questions through a series of thought experiments that transform theoretical questions whether artificial intelligence is possible into legal questions such as, "Could an artificial intelligence serve as a trustee?"'[47] He suggests that translating

---

[42] See Goodall, J., and Wise, S. M., 'Are Chimpanzees entitled to fundamental legal Rights?', 3 *Animal L.* 61, 1997 available at www.nabr.org/AnimalLaw/Articles/Goodall_Wise_AreChimpanzeesEntitled1997.pdf. For an historical overview of animal rights in continental and common law systems, see Epstein, R. A., 'Animals as Objects, or Subjects, of Rights', in: Olin. J. M. *Law & Economics Working Paper* No. 171, 2002, available at www.law.uchicago.edu/Lawecon/WkngPprs_151-175/171.rae.animals.pdf.

[43] *Infra*, s. 1.4.3.3.

[44] Note that much of this literature takes a common-law perspective, but the arguments are usually sufficiently general to be valid for continental legal traditions as well.

Within the scope of this article, we cannot go into all literature written on the subject. We refer interested readers to additional views expressed in, *inter alia*, D. Bourcier (2001), 'De l'intelligence artificielle à la *personne virtuelle*: émergence d'une entité juridique?', 49 *Droit et Société* 847; Emily M. Weitzenboeck (2001), 'Electronic Agents and the Formation of Contracts', *9 International Journal of Law and Information Technology* 204; R. George Wright (2001), 'The Pale Cast of Thought: on the Legal Status of Sophisticated Androids', 25 *Legal Stud. F.* 297; Chopra, S. and White, L. (2004), 'Artificial Agents - Personhood in Law and Philosophy', in: *Proceedings of the European Conference on Artificial Intelligence*, IOS Press, 635-639; W. Al-Majid (2007), 'Electronic Agents and Legal Personality: Time to Treat Them as Human Beings', *Proceedings of the 2007 Annual BILETA Conference, Hertfordshire, 16-17 April*, http://www.bileta.ac.uk/Document%20Library/1/Electronic%20Agents%20and%20Legal% (last accessed 17 March 2009); F. Andrade, P. Novais, J. Machado and J. Neves (2007), 'Contracting agents: legal personality and representation', 15 *Artif. Intell. Law* 357.

[45] Lawrence B. Solum (1992), 'Legal Personhood for Artificial Intelligences', 70 *N.C.L.Rev.* 1231.

[46] For a relevant discussion of subsequent paradigms in AI, see Varela, F. J., E. Thompson, et al. (1991), *The Embodied Mind. Cognitive Science and Human Experience*, Cambridge, MA: MIT, and Hayles, N. K. (1999), *How we became posthuman. Virtual bodies in cybernetics, literature, and informatics*, Chicago: University of Chicago Press.

[47] Solum, *supra* note 45 at 1232.

questions about AI in a concrete legal context will act as a pragmatic Occam's razor,[48] because the law allows us to detect the practical implications of providing legal personhood for smart technologies.

1.4.1.1 Personhood for non-humans: a legal fiction?

Referring to John Chipman Gray's *The Nature and Sources of the Law* of the beginning of the 20th century, Solum recounts the traditional idea that legal personhood for non-humans involves a fiction unless the entity can be said to have 'intelligence' and 'will'.[49] In order to avoid controversial terms like 'will' and 'intelligence', Solum investigates whether an AI could serve as a trustee (perform complex actions) or claim constitutional rights and liberties (assuming intentionality and consciousness).

Solum thus redefines the conditions for legal personhood in terms of the capacity to perform complex actions and/or the capacity to act intentionally and with (self-)consciousness.[50] The second capacity seems to comply with the traditional idea amongst many lawyers, philosophers and ethicists that personhood implies the capacity to act in a deliberate way. We should note, however, that legal personhood is often attributed to entities that do *not* qualify for such personhood (like ships, corporations etc.). Legal theory refers to this as a legal fiction: the law attributes personhood though in 'normal' life we would not think of the relevant entity as a person. Ironically, the traditional idea that legal personhood for non-humans is a legal fiction has been challenged by Tom Allen and Robin Widdison.[51] In fact they claim that in as far as contracts are initiated, negotiated and concluded by autonomous computers,[52] this attribution would imply a legal fiction if the legal consequences of these actions were attributed to the owners or users of these computers. In as far as they are not even aware of the contracts being concluded, it would be fictitious to pretend they concluded the contracts. This position is not contrary to Solum's. He argues for a pragmatic approach to legal personhood: for him the question of whether we need legal personhood is empirically dependent on the measure of independence of the artificial intelligence he discusses. Such independence depends on the capability to perform complex actions (reducing the need for human intervention) and – in the case of claiming constitutional rights and liberties – on the capability to have conscious intentions.

In the next section, we will discuss the question whether an AI can serve as a trustee (whether it has the capacity to perform complex actions), and in the following section we will discuss whether AIs can claim constitutional rights and liberties (assuming intentionality and consciousness). The discussion of AIs acting as a trustee is relevant for the question of granting a restricted form of legal personhood to computer agents in order to bridge the accountability gap in cases that do not depend on the attribution of guilt or wrongfulness. The

---

[48] Occam's razor is a 'principle stated by William of Ockham (1285–1347/49), a scholastic, that *Pluralitas non est ponenda sine necessitate*; "Plurality should not be posited without necessity." The principle gives precedence to simplicity; of two competing theories, the simplest explanation of an entity is to be preferred.' See 'Ockham's razor', *Encyclopædia Britannica* Online, http://www.britannica.com/EBchecked/topic/424706/Ockhams-razor (last accessed 9 March 2009). Solum here refers to this principle because he wants to avoid complicated metaphysical debates about what is 'intelligence', 'agency', 'personhood', etc.

[49] Solum, *supra* note 45 at 1238, footnote 26: Gray, J.C., *The Nature and Sources of the Law,* (ed. By Roland Gray in 1921, original publication in 1909). See also French, *supra* note 40 about what types of entities qualify for moral and legal personhood.

[50] We note that Solum does not discriminate between consciousness and self-consciousness, often using the term 'consciousness' to refer to self-consciousness. As explained *supra*, s. 1.3.2 on autonomic behavior and autonomous action, we think this to be a crucial difference.

[51] *Infra*, s. 1.4.2.1.

[52] Allen and Widdison speak of 'autonomous' computers, whereas we would qualify these computers as autonomic devices; *Cf. supra*, s. 1.3.2.

discussion of AIs claiming constitutional rights and liberties is relevant for the question of granting full legal personhood, bridging the accountability gap in the case of criminal liability for harm caused, and facing the issue of whether this implies that these entities have fundamental (post)human rights.

1.4.1.2 Acting as a trustee: the capacity to perform complex actions

To test whether an AI could perform the type of complex actions that are required for legal personhood, Solum describes three stages of involving an expert system in the management of a trust.[53] The first stage concerns an expert system that advises a human trustee to invest in publicly traded stocks, to pay the monthly bills to the beneficiary and to fill in the forms for tax returns. The actual performance of day-to-day tasks is largely automated but the final decisions are all taken by the human trustee. The second stage concerns an expert system that begins to outperform the human trustee as an investor, which leads to the settlor deciding to include instructions in the terms of the trust to the effect that the human trustee must follow the advice of the expert system. The role of the human trustee diminishes and the number of trusts that can be handled by the expert system increases exponentially. All routine interventions of the human trustee (e.g., in the case she is frequently sued by a beneficiary) are taken over by the expert system, producing letters that need only a signature of the human trustee. The third stage begins when the settlor decides to 'do away with the human trustee' because he wishes to save money or does not trust the human who may succumb to the temptation to embezzle funds. Now, who owns the expert system? If it were a legal person it could claim an ownership right to the hardware and software that allow it to operate, but since expert programs have no legal subjectivity under contemporary law, the hardware and software are probably owned by another legal person, e.g., a company. Having introduced these three stages, Solum raises the legal question of: 'whether an AI can become a legal person and serve as a trustee'. For the sake of the argument, he assumes that the trust does not raise complex moral or aesthetic issues and that it gives the trustee very little discretion. He also assumes that the expert system can make sound investments, take care of automatic payments and recognize events such as the death of the beneficiary which requires a change of action. He then pins down the issue to the question of 'whether the AI is competent to administer the trust'. Against the idea that an AI could serve as a trustee, he anticipates two objections: (1) the responsibility objection and (2) the judgment objection.

The responsibility objection

The thrust of this objection is that the expert system could not compensate the trust and cannot be punished if it violates legal obligations like the exercise of reasonable skill and care in investing the trusts assets or if the expert system embezzles trust assets. Presently the manufacturer of the system can be held liable on the basis of product liability. Can we imagine the system itself to be held liable? How could it compensate for damages? Solum suggests the system could be insured, but admits that civil liability for intentional wrongdoing or criminal liability are hard to imagine in the case of an expert system. In response to the

[53] A trust is a legal instrument in common law. It is 'defined as '"a fiduciary relationship with respect to property subjecting the person by whom the title to property is held to equitable duties to deal with the property for the benefit of another person, which arises as a result of a manifestation of an intention to create it." *Restatement (Second) of Trusts* § 2 (1959). The trustee is the legal person who administers the trust – invests trust assets, and so forth. The beneficiary is the person for whom the trust is maintained, for example, the person who receives income from the trust. The settlor is the person who establishes the trust. The terms of the trust are the directives to the trustee in the document or instrument creating the trust' (Solum, *supra* note 45 at 1240n). In the continental legal tradition the function of trust is taken by several institutes. The one Solum refers to might be closest to that of a foundation.

objection, Solum discusses the reasons for punishment. He argues that if deterrence is the reason for punishment one could claim that since expert systems can be designed in a way that makes it incapable of stealing or embezzling, there is simply no need for punishment. On the other hand, if desert or retribution is the reason for punishment, one could claim that non-human entities are not capable of the moral judgment that is required if one is to attribute desert and retribution. Finally, if punishment is a learning process, Solum cannot imagine which punitive action could communicate censure to the program. He thus concludes that regarding civil liability legal personhood for an expert system could work out for as far as the system can be insured for its liability. As to criminal liability or civil liability for intentional wrongdoing, he finds that liability is hard to imagine.

The judgment objection

The thrust of this objection is that an expert system will always consist of a – possibly – complex system of rules, which does not allow the system to make judgments in the sense of exercising discretion. The objection is played out in three versions. First, the argument is that an AI cannot cope with a change of legally relevant circumstances; second, it cannot make the moral choices it may encounter; and third, it cannot make some of the legal choices it will face. In all three versions, the problem is that – even in the case of parallel distributed algorithms – an expert system cannot but follow rules. As to the first argument, expert systems seem to lack the kind of common sense needed to solve unexpected problems, as to the second argument they seem to lack the 'sense of fairness' that is warranted when unexpected circumstances require one to overrule the letter of a rule in order to serve its purpose, and as to the third argument they seem to lack the ability to take the necessary action if called to account in a court of law. Solum concludes that AIs presently do not have the capacity to perform the duties of a trustee, especially in the case of unexpected circumstances affecting the trust. He raises the question whether a more limited form of legal personhood could be designed, allowing an AI to serve as a limited purpose trustee and/or for simple trusts whose operation can be fully automatic. In that case the terms of the trust will need to specify a human take-over whenever unanticipated circumstances rule out automatic behavior. We note that Solum seems to restrict himself here to *automatic* devices. Where *autonomic* computing is concerned, it seems that responsiveness to changed circumstances is part of its definition: even if the system cannot but follow rules, it is supposed to be capable of adjusting the rules that determine its performance. The first objection may thus fail in the case of autonomic devices. As to the third objection this also applies to corporations and funds to which legal personhood has been attributed. This leaves the second objection as the only real objection with regard to autonomic computer agents.

Limited personhood: who is the real trustee?

In the case of limited personhood the terms of the trust could stipulate that a natural person should take over in case discretionary judgment, requiring normative evaluation, is needed. This raises the question of who is the real trustee here. Why attribute limited personhood if in the end the real decisions have to be taken by a delegated or substituted natural person? This objection can be read in two ways: first, one can understand it as meaning that it is an essential quality of a trustee to have the ability to make discretionary decisions; second, one can read it as implying that the ability to make such decisions is just a practical corollary of trusteeship – someone has to decide at some point on unforeseen issues. Solum rejects the first reading as unnecessarily 'essentialist'. The second reading, however, allows Solum to conclude that the added value of providing a form of legal personhood to a non-human can be found in the fact that most decisions are routine rather than discretionary, and it may seldom be necessary to go back to a natural person for a discretionary decision, thus making the AI

function as trustee for most practical purposes. Therefore, there is added value in economic terms: it may be cheaper to employ an AI as a trustee whenever routine handling of affairs suffices, while the risk that an AI embezzles or frauds is practically non-existent, thus diminishing losses due to such risks.[54]

1.4.1.3 Posthuman rights and liberties: the capacity for intentional action and (self-)consciousness

Next, Solum discusses whether an AI could claim constitutional rights and liberties, an issue closely related to the question of autonomous action and agency in traditional ethical and philosophical discourse. We will follow his argument as it may clarify some of the issues raised in the previous sections. We should keep in mind that Solum was writing at a moment when autonomic computing was hardly dreamt of, whereas today it looms just across the horizon. The scenario on which Solum's question builds is one of relatively independent artificial agents that function as a kind of human-machine-interface (HMI) that locates relevant information for a human person, for instance in her professional life. Considering their computing power, they are capable of intelligent mining of a knowledge domain and of knowledge management far beyond the reach of the human brain. As Solum writes, these HMIs seem to have a 'mind of their own'.[55] He then advances the idea that at some point in time these independent AIs could claim constitutional rights like free speech (1st Amendment US Constitution) and the right not to be subject to involuntary servitude (13th Amendment US Constitution), meaning they would resist being owned by another person.

The question Solum wishes to raise is: 'whether we ought to give an AI constitutional rights, in order to protect its personhood for the AI's own sake'.[56] We rephrase this question as the issue of whether computer agents would qualify for a claim to what we will call posthuman rights and liberties, suggesting that at some point fundamental human rights like privacy, due process, and bodily integrity may be claimed by and/or attributed to non-human agents. By calling them post*human* rights and liberties we refer to the existing category of human rights and liberties; by calling them *post*human we acknowledge that they would apply for instance to non-biological machines, cyborgs or synthetic biological entities, while acknowledging that this may require us to rethink the meaning of existing human rights.[57] Solum again raises three kinds of objections: first, one could argue that only natural persons qualify for constitutional rights of personhood, second, one could insist that AIs lack some critical aspect of personhood, and third, one could suggest that since AIs are human creations, they can never be more than human property. Though it may seem cumbersome to investigate these objections, we nevertheless take time to explain them, as well as Solum's response. We think that an adequate answer to the question of whether computer agents qualify for legal personhood will benefit from a serious consideration of these objections. To be sure, there may be more aspects that affect the question of whether full legal personhood can be attributed – we should point out that Solum's points regard neither necessary nor sufficient conditions for full legal personhood – but any discussion of this matter must at least address the objections that Solum has put on the agenda.

The natural person objection

---

[54] Solum, *supra* note 45 at 1253-54.
[55] Solum, *supra* note 45 at 1256.
[56] *Ibid.* at 1258.
[57] At this point we do not move into the discussion of whether such post-human rights concern only first-generation (individual) rights or also second-generation (social) rights. We rather assume that a society of post-humans may require further generations of fundamental rights.

Though one could claim that some constitutional rights should be restricted to human persons, we must acknowledge that specific constitutional rights (like the Equal Protection Clause and the Due Process Clause in the US Bill of Rights) already apply to non-human legal persons, while corporations also have a right to freedom of expression. The objection, however, maintains that in those cases the non-human legal person is no more than a place-holder for the rights of natural persons. A more fundamental argument against constitutional rights for non-humans holds that the concept of person is intrinsically linked to humans. The idea is that since non-humans do not share our biological constitution, they cannot be conceptualized as persons. Solum counters this point by arguing that the fact that today we cannot imagine non-humans to qualify for personhood, does not imply that, in the future, AIs could not develop into non-biological entities that are intelligent, conscious and feeling in ways that *change our very concept of personhood*. We add that the advent of cyborgs and synthetic biology blurs the border between biological and non-biological entities. Cyborgs, defined as humans enhanced with implants that – for instance - change brain functioning, seem to introduce a continuum between non-biological robots and human-machine hybrids. While Dreyfus had a point against the first generation of AI 'believers', when he emphasized that cognition is always embodied cognition (a point well taken, resulting in 'embodied computing' paradigms), others stress that embodiment is not necessarily biological embodiment.[58] Following Solum's argument we may expect non-biological embodiment as well as cyborg embodiments to provoke novel conceptions of personhood.[59] Finally, socio-biological and utilitarian arguments that it is not in our interest to grant constitutional personhood to AIs because they may take over, seem to miss the point: they assume that moral obligations are only in play between humans and they ignore the fact that if AIs could take over this would certainly not depend on us granting them any rights. If we build machines that develop intelligence, consciousness and feeling – Solum seems to suggest – we take the risk of entering a new society of both human and non-human persons.

The missing-something objection

This argument basically evolves as follows: something (the soul, consciousness, intentionality, feelings, interests, free wills) is essential for personhood.[60] As no AI can have this 'something', the simple fact that a computer could *simulate* having this something does not mean it actually does have it. Since having this 'something' determines humans as persons, non-humans cannot be persons.

Regarding the argument of non-humans not having a **soul**, Solum explains that in as far as this is a theological argument it cannot determine the attribution of legal personhood: in a pluralist society legal or political arguments need to be based on public reason, i.e., reasons

---

[58] Dreyfus, H.L. (1992), *What Computers Still Can't Do. A Critique of Artificial Reason*, Cambridge, MA: MIT Press; Ihde, D. (2002), *Bodies in Technology*, Minneapolis/London: University of Minnesota Press; Jos de Mul (2002), *Transhumanism. The convergence of evolution, humanism, and information technology*, available at http://www.filosofie-in-bedrijf.com/uploadedFiles/Brand_information/Perspectives/Transhumanism%20by%20Jos%20de%20Mul.pdf (last accessed 9 March 2009).

[59] *Cf.* the cyborg sense of self, described in Warwick, K. (2002), *I, Cyborg*, London: Century, and Warwick, K. (2009), 'Implants and cyborgs: the environment and the self', in: Koops, B.J., M. Hildebrandt and K. De Vries (eds.), *D7.14b: Idem-Identity and Ipse-Identity in Profiling Practices*, FIDIS deliverable, available at http://www.fidis.net/resources/deliverables/ (last accessed 28 April 2009).

[60] This issue of what this 'something' is has been debated ever since the AI community began to take serious the objection that computation and manipulation of symbols cannot explain human self-consciousness. See Varela et al. (1991) and Hayles (1999), *supra* note 46 and Dreyfus (1992), *supra* note 58. For interesting overviews, see Holland (2004) and Woodruff Smith and Thomasson (2005) *supra* notes 34 and 35, and, seminally, Graubard, S.R., ed. (1988), *The Artificial Intelligence Debate: False Starts, Real Foundations*, MIT Press.

that people from all different religious or non-religious beliefs can accept. In as far as the argument builds on a Cartesian duality between material causality and mental freedom, he finds it inextricably wound up in the pitfalls of an untenable dualism. Regarding the argument of non-humans not being capable of possessing **consciousness**, Solum explains that if AIs are in fact incapable of having self-consciousness they would not be capable of experiencing their own life as good or evil, nor could they develop ends. According to Solum ends or goals are a precondition for being a right-holder. However, the question of whether AIs are capable of developing self-consciousness is an empirical question. Though at this moment consciousness seems restricted to biological beings, this in itself does not preclude the possibility of non-biological consciousness. The empirical question is complicated because a computer may simulate having consciousness, as a strategy to successfully claim constitutional rights, but this still does not rule out altogether that AIs may one day convince us of their self-consciousness. We would suggest that if AIs could in fact *simulate* consciousness as a *strategy* to claim constitutional rights, one would be tempted to infer that they have at least some kind of consciousness. Regarding the argument of non-humans not being capable of possessing **intentionality**, Solum explains that intentionality refers to 'meaning'. Just like a thermostat may seem to 'know' whether it is too hot or too cold in a room, an AI may seem to know which stocks to buy. However this 'knowledge' does not imply even the faintest idea of the *meaning* of hot and cold or expensive and cheap. So far, AIs seems to excel in *syntactics*, without having a clue as to the *semantics* of what they are 'doing'. The argument would be that as long as computers cannot give 'meaning' to their own life, it makes no sense to attribute constitutional rights.[61] However, like in the case of consciousness, Solum argues that we cannot preclude AIs from developing meaning. Regarding the argument that non-humans cannot possess **feelings**, Solum discusses the experience of emotions, desires, pleasures and pain. Though he has some doubts about whether personhood depends on having feelings, he moves on to discuss 'what if' emotions, pain and pleasure were to be essential for the attribution of personhood.[62] The argument then develops similarly as in the case of consciousness and intentionality: it may be that having feelings depends on our biological constitution, but it may also be the case that in the future AIs will develop feelings, though these feelings will be embodied differently from ours. In that case, he sees no reason to deny personhood for an AI. Regarding the argument that non-humans cannot possess **interests,** defined as an interest in the good life, Solum discusses the utilitarian idea that the good life is defined as maximizing pleasures and minimizing pain. In that case, the question of whether they can have interests equates with the question of whether they have feelings. However, if one takes a more objective and public perspective on interests, like John Finnis does for example, the question is whether an AI can flourish by including goods such as 'life, knowledge, play, aesthetic experience, friendship, practical reasonableness, and religion'.[63] Solum contends that even if AIs will not have a 'life' in the biological sense of the word, they might lay claim to a life in which goods like knowledge, play, friendship etc. can be realized.

---

[61] Giving meaning to one's life refers to the fact that humans are 'symbolic' animals, whose cognition is mediated by natural language. This language has semantical, syntactical and pragmatic dimensions, whereas computer language is limited to a syntactical and (in the case of decision systems) pragmatic dimension. A computer agent does not 'understand' what it is 'doing' in the symbolic terms associated with the notion of 'giving meaning to one's life'.

[62] Solum *supra* note 45 at 1270 seems to agree with Kant that all rational beings qualify for personhood, irrespective of them having feelings. He also refers to Aaron Sloman's argument that any system with multiple goals requires a control system, with emotions achieving just that in the case of human beings. This seems to be confirmed by research demonstrating that intelligent people with brain-damage that reduces their capacity to be emotional can give multiple arguments for any course of action but remain incapable of making decisions.

[63] Solum *supra* note 45 at 1271n, referring to John Finnis, *Natural Law and Natural Rights*, 1980, at 85-90.

Moreover, if living in a pluralist society implies that we accept alternative conceptions of the good life, we should make room for radically different ways of conceptualizing the good life, for which the attribution of personhood is in fact a precondition. Regarding the argument that non-humans cannot possess **free wills**,[64] being the precondition for autonomous action, Solum explains that in as far as AIs are merely an instrument to execute the free will of a human being, they could not qualify for personhood. The argument thus focuses on the issue of whether an AI could ever 'act' beyond the instructions (the program) of the human that designed it. Are the actions of an AI entirely mechanical, or could we imagine them as capable of conscious deliberation, reasoning, and planning?[65] Again, this is an empirical question: we cannot preclude the possibility that AIs will develop 'a mind of their own', capable of conscious reflection, deliberation and planning. The fact that we could use a mechanical device to overrule an AI that does not obey our instructions, would – in itself – not be an argument against the attribution of personhood. It could be that this device is used precisely because the AI has developed its own reasons and plans; such a device would be like the discipline or punishment we exercise over other human beings, depriving them of the exercise of their free will rather than assuming they do not have one. We like to add that autonomic computing implies that the relevant digital agents 'act' beyond the instructions or algorithms of their human designer or user. This does however not imply that they have self-consciousness and plan or deliberate consciously about different courses of action. We must discriminate between autonomic and autonomous action. Both imply creative and partly unpredictable interactions, but autonomic action does not imply self-consciousness or the capacity for reflection that is at stake in autonomous action.

Summing up, in the case of souls and interests, Solum argues that the pluralism of our society should prevent us from imposing our own conceptions on spiritual matters or the good life on emerging AIs. In the case of consciousness, intentionality, feelings and free will, we should let empirical evidence decide the matter. As to the latter, Solum turns to the objection that we may apply the Turing test to AIs and find that they behave as persons,[66] while in fact they are only **simulating**.[67] He points out that to make the distinction between the simulation

---

[64] Within the cognitive sciences some authors claim that 'free will' is an illusion anyway; see, e.g., Wegner, D. M. (2002), *The illusion of conscious will*, Cambridge, MA: MIT Press. Though it seems obvious that much of our behavior is autonomic, this does not imply that there is no room for deliberate(d) action. In fact it seems obvious that our tacit forms of behavior are often learned behavior, initiated by conscious deliberation with others. The discussion whether and to what extent free will is an illusion relates to the issue of determinism and voluntarism. Rather than embracing either of these extremes we will presume that human action is underdetermined due to its mediation by natural language that allows us to externalize our thoughts and reflect upon them. This does not imply a mentalistic free will, nor a physicalist determinism; it remains agnostic as to the actual extent of our freedom. It must be clear, however that to hold a person accountable on the basis of culpable and wrongful action, a court must decide that this person has a measure of freedom to act otherwise. If this is denied, the attribution of rights and liberties in law makes no sense anyway.

[65] Solum *supra* note 45 at 1273 refers to the idea that human actions are not caused, meaning that the free will is not subject to the laws of causation. He rejects this as an implausible proposition, suggesting that 'an action is free if it is caused in the right way – through conscious reasoning and deliberation'.

[66] The so-called Turing test was proposed by Alan Turing (1950) in his classic paper 'Computer Machinery and Intelligence', 59 *Mind* 433. Turing suggests that if a person sitting behind a computer screen cannot detect the difference between the answers generated by a digital computer and those of a human person, this proves that the computer can think. *Cf.* Searle's classic refutation in Searle (1990), 'Is the Brain's Mind a Computer Program?', 262 *Scientific American* 26.

[67] This is Searle's Chinese Room argument, discussed by Solum *supra* note 45 at 1236-1238. It concerns the fact that a computer makes its inferences on the basis of syntactical correlations, without any semantic reference. Though the inferences could allow the computer to pass the Turing test, this would merely indicate that the computer can simulate a person without actually being one. On the question to what extent a Turing test should be relevant as evidence of personhood in a court of law, see *idem* at 1280.

of a person and the actual being of a person, behavioral evidence of great syntactical abilities would perhaps not be sufficient.

The objection that AIs should be property
This objection refers to Locke's proposition that artifacts that are the product of human labor are the property of those who made them.[68] For Locke, a human being is not 'made' by his parents but by God, implying that a parent does not have ultimate control over his children. Solum rejects this theological argument and asserts that we believe in personhood for all human beings, even if they are 'made' by their parents. The question is whether the fact that human beings are made 'naturally' while AIs are made 'artificially' should make a difference here. Solum believes the argument does not really add to the debate: whether an AI should be granted constitutional rights depends on it being a person and in as far as this is the case an AI should not be owned by another person. Moreover, even if AIs come into the world as the property of their makers, like slaves, they can emancipate and become 'free' persons. Or, as artificial slaves, 'they might still be entitled to some measure of due process and dignity'.[69]

1.4.1.4 Conclusion

Solum basically concludes that one could employ an intelligent non-human system as a trustee, attributing it a measure of legal personhood that fits the restricted capabilities of a system that is capable of autonomic decision-making even if it does not 'understand' the meaning of its decisions and does not have a goal in life (and does not really have a 'life' in our sense of the word). We think that his arguments are valid for computer agents like avatars, robots, and software programs that function in a sufficiently autonomic way. In the following section, we will take a more detailed look into the legal intricacies of the validity of and liability for contracting by (means of) computer agents. Building on Solum's discussion of constitutional rights for AIs, we think that as long as the behavior of computer agents is ultimately syntactical, based on correlations that have no meaning because the system has no consciousness of the world around it, we cannot grant posthuman rights and liberties that presume the capability to reflect upon one's actions, initiate intentional action and take responsibility. For the same reasons, it does not make sense to hold contemporary computer agents liable on the basis of culpable and wrongful action. We will return to this issue below, especially in the discussion of Teubner's position.[70]

1.4.2 Contracting and limited personhood

1.4.2.1 Allen and Widdison (1996)

In 1996, Tom Allen and Robin Widdison investigated the issue of the legal implications of digital contracting by computer systems that operate *not just automatically* but *autonomously*.[71] They define autonomous machines as those that first, can learn through experience, second, modify the instructions in their own programs, and third, devise new instructions.[72] This sounds very much like what IBM has recently coined 'autonomic computing', which is defined as: self-management, self-configuration, self-optimization, self-

[68] Solum *supra* note 45 at 1276 (footnote 159) referring to John Locke, *Two Treatises of Government* §§ 25-51, at 285-302 (Peter Laslett, ed. 1988/1690).
[69] Solum *supra* note 45 at 1279.
[70] See *infra*, s. 1.4.3.2.
[71] T. Allen and R. Widdison (1996), 'Can Computers Make Contracts?', 9 *Harvard Journal of Law & Technology* 26.
[72] *Ibid.* at 26.

healing, and self-protection.[73] Allen and Widdison anticipate that what we shall call autonomic agents could be used for computer-generated business-to-business transactions on the internet, especially for one-off transactions that are not performed in the framework of predetermined trading relationships. They envisage that such 'on the spot' trading would encourage 'just-in-time' ordering and stock control. They argue for adequate legal protection of such transactions, to ensure that the legal consequences can be effected, for instance when it is unclear who is 'behind' such autonomically concluded contracts. One way to provide a legal infrastructure that generates reliable agreements could be to register autonomic electronic agents that initiate, negotiate and conclude contracts for a company, as agents for the company in a public register. This would enable contracting parties to locate the responsible (legal) person 'behind' the agent.

Allen and Widdison discuss four ways of dealing with autonomic agents that initiate, negotiate and conclude contracts: first, modifying contract doctrine; second, seeing the computer as a tool of communication; third, in the traditional analysis, denying validity to transactions generated by autonomous computers; or fourth, conferring legal personality to computers. We note that their usage of the term 'computers' seems a bit awkward, as they are basically referring to interconnected systems rather than single computers. For this reason we will discuss their suggestions as relating to autonomic computer agents, which will typically be interconnected systems.

Modifying contract doctrine

As to the first option, the authors find that relaxing the requirement of intentionality in contract-making could solve the problem of computer-generated contracts: 'the court would hold that the human trader's generalized and indirect intention to be bound by computer-generated agreements is sufficient to render the agreements legally binding'.[74] This would fit well with the fact that the 'real' intentions of a contracting party will always remain virtual: they will be 'read' into the concrete interactions that lead others to trust the party's intention. We should however remember that the human parties that are thus bound by the contract may not know the exact terms of the contract and often not even be aware of the contract being concluded. The entire legal framework of offer and acceptance is replaced by machine-to-machine communication.[75]

The computer as a tool of communication

As to the second option the authors argue – as already indicated above[76] – that in the case of autonomic agents this approach creates a legal fiction: the agents are regarded as if they are a mere instrument in the hands of the contracting parties, while in fact they interact autonomically. They remark that unexpected and unreasonable contractual obligations could arise by which the parties would nevertheless be bound. If the agents could be regarded as legal agents, courts could use the legal doctrine of actual and ostensible agency to mitigate the legal obligations.

---

[73] See J.O. Kephart and D.M. Chess (2003), 'The Vision of Autonomic Computing', 36 *Computer* (1) 41.

[74] Allen and Widdison, *supra* note 71 at 44.

[75] This is the difference between what has been coined as 'Ambient Law' in M. Hildebrandt & B.J. Koops (eds.) (2007), *D7.9: A Vision of Ambient Law*, FIDIS Deliverable, October 2007, available at http://www.fidis.net, and M. Hildebrandt (2008), 'A Vision of Ambient Law', in: Brownsword and Yeung, eds, Regulating Technologies, Oxford: Hart Publishing, 175-191. Ambient Law would imply that a legal norm is articulated into a technology, which means that the legislator is aware of the affordances of the technology and also requires that if legal consequences are attributed to the violation of a norm, this is made contestable in a court of law. Replacing a legal by a technological framework is something altogether different, and could easily enforce norms in a way that places them outside the reach of the legal and constitutional framework.

[76] *Supra*, s. 1.4.1.1.

Actual agency is defined as: "the agency that exists when an agent is in fact employed by a principal".[77] Ostensible or apparent agency is defined as:[78]

> agency by estoppel: an agency that is not created as an actual agency by a principal and an agent but that is imposed by law when a principal acts in such a way as to lead a third party to reasonably believe that another is the principal's agent and the third party is injured by relying on and acting in accordance with that belief[.] A principal has a duty to correct a third party's mistaken belief in an agent's authority to act on the principal's behalf. If the principal could have corrected the misunderstanding but failed to do so, he or she is estopped from denying the existence of the agency and is bound by the agent's acts in dealing with the third party.

We should note that for ostensible agency an action is required by the principal; she cannot be bound to a third party if there is no action of the principal that leads a third party to reasonably believe that the alleged agent is an actual agent. Should the principal, however, decide to 'pull' the contract towards himself by means of ratification, he will be bound by the contract.

Another important part of the law of agency that is relevant here is the doctrine of disclosed and undisclosed agency:[79]

> Continental European laws restrict the application of agency rules to cases where the agent acts openly in another's name. Thus, French jurists infer from article 1984 of their Civil Code, according to which agency is the act of the agent *pour le mandant et en son nom* ("for and on behalf of the principal"), the negative conclusion that in case an agent does not disclose that he is acting as an agent for a principal, the consequences touch only the "agent" himself. The hidden principal is not concerned by the effects of the transaction at all. Section 164 of the West [sic] German Civil Code expressly provides that "an agent, who acts without disclosing the fact that he is acting as agent, is the only one to acquire any rights and is exclusively personally liable."
>
> In contrast to the continental view, when an agent contracts in his own name without disclosing his principal, the common law allows the undisclosed principal under certain conditions to sue or be sued by the third party. Such conditions include that the agent had power to make the contract and that the parties eventually learn their respective identities. This wider concept of agency has no counterpart in continental legal tradition.
>
> The use of this basic doctrine in the common-law countries gives rise to questions regarding the identity of the undisclosed principal, the election of remedies that must be made by the third party, the extent of the respective liabilities, the right of the third party to setoff (the amount of its own damages from any sum that might be awarded it), etc. A solution to these conflicts of interests must in final analysis rest upon an evaluation of the extent to which the relationship between the undisclosed principal and the agent should influence the contract made by the agent with a third party.

The categories of disclosed and undisclosed agency seems highly relevant for our subject, and we should take into account what it affords in the case of attributing legal personhood to electronic agents or e.g. multi-agent-systems (MASs).

Denying validity to transactions generated by autonomous computers

As to the third option the authors point out that as current doctrine demands human intention, the actions of autonomic digital agents could not lead to a valid contract. By not relaxing this requirement (as under the first option) human parties would not be obligated by the contracts concluded by their autonomic agents. The authors indicate that the enforceability of an automatically generated contract would become dependent upon whether the agent was an autonomic agent, while in fact this may not always be apparent to the other party. This would stifle commercial enterprise, in their opinion.

---

[77] See for definitions of agency and more clarification: http://dictionary.getlegal.com/agency (last accessed 9 March 2009).

[78] *Ibid.*

[79] Quoting from 'Agency', in *Encyclopædia Britannica* (2008), Encyclopædia Britannica Online, http://www.britannica.com/EBchecked/topic/8976/agency (last accessed 9 March 2009).

Granting legal personhood
As to the fourth option the authors investigate the moral entitlement, the social reality and the legal expediency of legal personhood for autonomic agents. They agree with Solum[80] that a moral entitlement to legal personhood would depend on them developing self-consciousness. However, while they agree that at present no sign of such self-consciousness has emerged, they find that the legal system could still recognize the social fact of the independent actions of autonomic digital agents. Referring to Teubner, they suggest that it makes sense to grant legal personhood to entities that are capable of what we call autonomic action. The point is not whether an agent *understands* the meaning of its actions (which would require consciousness and allow for autonomous actions). The point is only that since it is capable of developing a trading strategy of its own, it makes sense to make the agent responsible for such independent action. The legal expediency of granting legal personhood resides in allowing the agent to act as a legal agent (which is not possible for an entity without legal personhood), and to allow a contracting party to identify the digital agent as the legal agent of a specific company. They propose for companies to register their digital autonomic agents in a public register, stating the competence and limitation of liability.

1.4.2.2 Wettig and Zehendner (2003/2004)

Like Allen and Widdison, Wettig and Zehendner have discussed the legal implications of contracting by electronic agents.[81] Their analysis is much in line with Allen and Widdison's, but because it is based on continental law and German legal doctrine, it is useful to describe their argument here in addition to the previous section. They note that, in contrast to conventional software, electronic agents have, to greater or lesser extents, characteristics like reactivity, proactivity, adaptive behavior, mobility, and autonomy (which they define as 'the ability to operate without the direct intervention of humans or others, and (…) some kind of control over their action and internal state').[82]

Noting that declaration of intent (DOI) is a key factor in determining the legal status of a contract, Wettig and Zehendner distinguish between three forms of ICT-related declarations of intent: electronic DOI, where the intent is communicated by electronic means, automated DOI, where the intent are mechanically produced with the help of a computer program, and computer declaration (*Computererklärung*), where the declaration of intent is electronically produced in a completely automatic way, without being directly influenced by human action. The latter form can be seen as the declaration of what we call autonomic agents. These declarations are usually seen as a DOI of their user, comparing the agent to a vending machine (selling automatically to anyone who happens to use the machine) or a working tool (where the declaration functions as a signature in blank, without the user knowing the exact future contents of the contract).[83]

The issue now is whether contracting activities by electronic agents can, also in the future when the agents become increasingly autonomic, still be treated as 'computer declarations' under German law, which is the traditional approach of ascribing the intent of an electronic agent to a user. Particularly for mobile agents, with increasing spatial distance between principal and agent, this approach becomes troublesome, as the principal seems to have less

---

[80] *Supra*, s. 1.4.1.
[81] Wettig and Zehendner, *supra* note 14; S. Wettig and E. Zehendner (2003), 'The Electronic Agent: A Legal Personality under German Law?', in: Oskamp, A. and Weitzenböck, E., eds., *Proceedings of the Law and Electronic Agents workshop (LEA '03)* 97-112.
[82] Wettig and Zehendner (2003), *supra* note 81.
[83] Wettig and Zehendner, *supra* note 14 at 120-122.

direct influence over the agent. In a more modern approach, authors have suggested various analogies to ascribe legal personhood to electronic agents to solve this 'distance problem'. In this approach, three legal doctrines are applied to interpret electronic agents as having some form of legal personhood under existing law. First, the agent could be a representative (*Stellvertreter*) who declares his own intent with mandate of the principal; the problems here are that a representative needs to be a legal subject himself, which is problematic for electronic agents, and particularly that in case of false representation, the electronic agent is liable (legal doctrine supposing that the agent, not the alleged principal, is the contracting party) but cannot pay up by itself. Second, the agent could be a messenger (*Bote*) conveying the DOI of its principal; the problem here is that (autonomic) electronic agents do more than just transport messages: they influence the terms of the contract and are therefore not mere messengers. Third, the electronic agent could function as a minor, with limited capacity to contract itself (*beschränkt Geschäftsfähiger*); however, contrary to minors who contract on their own behalf (e.g. buying ice-cream), electronic agents contract on their principal's behalf, and the regulations for contracting by minors are therefore ill-suited.[84]

Since both the traditional and the modern approach are flawed to conveniently accommodate contracting by electronic agents under existing law, Wettig and Zehendner propose a 'progressive approach' of changing the law: granting legal personhood, under certain conditions, to electronic agents. Besides the natural person and the legal person, the electronic person (e-Person) could be created. Following up on Allen and Widdison's suggestion, the authors propose that companies register their autonomic electronic agents in a public register, stating the competence and limitation of liability. 'The result would be a kind of agent with limited liability (Ltd. Agent).' This has the advantage for the party contracting with the agent that they do not always have to trace back the mobile agent to its distant principal, but can check the agent's solvency in the register. For the owner of the agent, the advantage is that they can limit the agent's liability, and thus control the risk of using an autonomic agent over whose actions the owner has relatively little direct influence. Introducing this restricted form of legal personhood for electronic agents and a register with limited liability does not preclude users from applying unregistered agents, but for such agents that lack the restricted legal personhood, the actions would always be attributed to the owner.[85] Of course, changing the law in this way is only possible if an adequate definition can be given of electronic agents that are to have a claim to restricted legal personhood; Wettig and Zehendner point to the definitions of electronic agents in UETA, UCITA and the Canadian UECA[86] as a good starting point for coining an acceptable definition.[87]

1.4.3 Accountability: towards full personhood?

Whereas the previous section discussed the ways in which Allen & Widdison and Wettig & Zehendner have followed up on the first line in Solum's analysis – new types of entities acting as a trustee – we now turn to literature that builds upon Solum's second line: to what

---

[84] *Ibid.* at 123-126. At 127, the authors briefly reject an alternative 'historic approach' of interpreting electronic agents as slaves under Roman law, where the slaves can contract without having legal capacity, and their actions being attributed to their master. Arguing that in current law, contractual capacity presumes legal capacity, Wettig and Zehendner thus dispose of the historic suggestion by E. Schweighofer (2001), 'Vorüberlegungen zu künstlichen Personen: autonome Roboter und intelligente Softwareagenten', in E. Schweighofer et al., eds, *Auf dem Weg zur ePerson. Schriftenreihe Rechtsinformatik*, Vol. 3, Verlag Österreich: Wien, 45–54.

[85] Wettig and Zehendner, *supra* note 14 at 127-129.

[86] See s. 2(6) Uniform Electronic Transaction Act (1999); s. 102(27) Uniform Computer Information Transaction Act (1999); part II, para. 19 Canadian Uniform Electronic Commerce Act.

[87] Wettig and Zehendner, *supra* note 14 at 129-131.

extent can new types of entities be held accountable for moral wrongs and attributed rights and duties?

1.4.3.1 Karnow (1996)

In 1996, Karnow investigated the issue of legal solutions for harm caused by distributed artificial intelligences. His major point is that, at this moment, we see emergent AIs that operate in the real world with decision programs, making 'decisions unforeseen by humans'.[88] These unforeseen – and sometimes unforeseeable – decisions will at some point cause damage or injury, and Karnow claims that this will lead to 'insuperable difficulties (…) posed by the traditional tort system's reliance on the essential element of causation'.[89] He explains that the complexity of digital systems 'connotes multiple interacting but independent elements' making it 'difficult, and sometimes impossible, to predict the sum state of the complex system'.[90] As to search machines, Karnow anticipates that even 'classic "expert" systems that mechanically apply a series of rules to well-defined fact patterns' (automatic agents, in our terms) will not be able to mine relevant information, due to the persistent and exponential information growth.[91] Instead, what he calls 'intelligent agent technology' will be 'responsible' for the searching of relevant databases, and for deciding on relevant actions to be taken. His reference to intelligent agent technology confirms Allen and Widdison's discussion of what we have called autonomic agents. Karnow claims that these agent systems are relatively unpredictable, stating that 'Fixing' these unpredictable systems to operate predictably will eviscerate and render them useless', suggesting that '[t]rue creativity and autonomy require that the program truly makes its own decisions, outside the bounds expressly contemplated by either the human designers or users.'[92]

The problem with such unpredictability is that it generates errors and faults, due to what Karnow calls 'pathological decisions'. And such decisions are not something we can resolve by writing better programs. On the contrary, Karnow claims that '[t]hese are not 'bugs' in the programs, but are part of their essence.'[93] He speaks of the fact that 'the long-term operation of complex systems entails a fundamental uncertainty' precisely in the kind of complex and unpredictable environments that require the input of autonomic agents.[94] As these agents are both mobile and distributed, they easily move outside the control of their user and it becomes difficult to attribute causality to either the physical person or company that is 'behind' the agent. But as these agents interact within a networked world, it becomes equally impossible to attribute causality to a single node within a network (as the node builds on connectivity) or to the network as a whole. One of the reasons for this is that such intelligent agents will often be polymorphous (difficult to identify as the same agent), while on top of that the boundaries of the network are dynamic, raising similar difficulties of identification.

---

88 C.E.A. Karnow (1996), 'Liability for Distributed Artificial Intelligences', 11 *Berkely Technology Law Journal* 147 at 148.

89 Karnow, *supra* note 88 at 148-149.

90 Karnow, *supra* note 88 at 149.

91 *Cf.* also Kallinikos, J. (2006), *The Consequences of Information. Institutional Implications of Technological Change*, Cheltenham, UK/ Northampton, MA: Edward Elgar.

92 Karnow, *supra* note 88 at 154,161. *Cf.* also G. Sartor (2002), 'Agents in Cyberlaw', Sartor, G. and Cevenini, C. (eds), *Proceedings of the workshop on the Law of Electronic Agents (LEA02)*, 2002, available at http://www.lea-online.net/publications/Sartor.pdf (accessed 8 May 2008): 'Note that the difficulty of anticipating the operations of the agent is not a remediable fault, but it is a necessary consequence of the very reason for using an agent: the need to approach complex environment by decentralizing knowledge acquisition, processing and use. If the user could forecast and predetermine the optimal behaviour in every circumstance, there would be no need to use an agent (or, at least, an intelligent agent).'

93 Karnow, *supra* note 88 at 161.

94 Karnow, *supra* note 88 at 162.

Liability in law requires causality: without a causal relationship, one simply cannot attribute liability. Even in the case of strict liability, which forsakes traditional requirements like intent or fault, negligence, recklessness, or other types of culpability, tort liability cannot do without 'proximate cause'. The concept of 'proximate cause' is a typically legal notion, used to discriminate between what Karnow calls 'cause in fact' (i.e., what continental lawyers would call the *conditio sine qua non*) and the legally relevant cause.[95] The idea is that any event in real life has a multiplicity of causes that overlap and intermingle: from distant in time and space to relatively nearby or even concurrent causes. To establish liability, one needs to single out an event that allows the imputation of responsibility for harm suffered, which already limits the domain of possibly relevant causes to human action (including omission or neglect), or at least to actions attributable to a legal person. To single out the relevant causation amongst the mass of causally relevant events, lawyers speak of the 'proximate cause', which is often equated with a cause that 'brought about' harm that was 'reasonably foreseeable'.[96] The idea is that the (natural or legal) person that could have foreseen the harm should have prevented it. For the same reason, someone who caused an accident in the sense of 'causation in fact' may be absolved in law from having caused the accident because of what is called an 'intervening' or 'superseding' cause that is deemed more relevant for the harm caused. Imagine that a person breaks the bike of a friend, which makes him liable for the damage done to the property of his friend. Not having the bike, his friend walks to the supermarket and gets hit by a car. Though breaking the bike is a 'cause in fact' of the accident, courts will probably consider the collision with the car to be an 'intervening cause'. Karnow rightly explains that what is 'reasonably foreseeable' depends on custom and common sense, meaning that in a fast changing environment like today's, 'reasonable foreseeability is a moving target'.[97] This keeps the legal system alert and responsive to societal developments.

However, Karnow then moves on to discuss causality in an environment with autonomic agents. His main point is that such an environment will come to a point where the attribution of legal causality (the establishment of proximate cause) does not make sense anymore. The reason for this is that autonomic agents, cooperating across a distributed network, will develop what he calls 'pathological decisions' next to routine and highly original, successful decisions. Such decisions are not always predictable, they are not a matter of preventable error or bugs in the system, but – as argued above – part and parcel of the intelligence of the network. Karnow basically warns that we cannot have our cake and eat it too: autonomic agents will solve problems we could not have solved ourselves, but this will also involve an 'unpredictable pathology'.[98] To attribute liability to any (human or non-human) node in the network, or even to the network itself, would create an arbitrary legal fiction that has no purpose in the law: since nobody could have foreseen this decision, nobody could have prevented it, so imputing causality or liability makes no sense. As Karnow explains:[99]

> 'The notion of "proximate" or "legal" causation implies a court's ability to select out on a case-by-case basis the 'responsible' causes. But where damage is done by an ensemble of concurrently active polymorphic intelligent agents, there is insufficient persistence of individual identifiable agencies to allow this form of discrimination.'

One way of dealing with this situation would be to ban autonomic agents altogether. One could imagine that the principle of precaution is at play here, requiring more research into the potential consequences of harm 'caused' by entities that cannot be held responsible before

[95] Karnow, *supra* note 88 at 176.
[96] Karnow, *supra* note 88 at 178.
[97] Karnow, *supra* note 88 at 181.
[98] Karnow, *supra* note 88 at 188.
[99] Karnow, *supra* note 88 at 191.

introducing a technology with irreversible consequences. Another option, chosen by Karnow, is to abolish legal liability in such a case and to seek a technological solution for what he deems to be a technological problem. Instead of hanging on to the traditional tort system and trying to control the uncontrollable, Karnow proposes a Turing Registry. This Registry would enlist certified autonomic agents that are insured against the risk of pathological decisions, meaning that even when no proximate cause can be established (thus excluding strict liability) the relevant agent is at least insured in order to compensate for damages. How this solves the difficulties of identification of a polymorphous agent consisting of a network with unstable boundaries is not altogether clear.

1.4.3.2 Teubner (2007)

In a provocative Max Weber Lecture at the European University Institute in Florence,[100] Gunther Teubner has argued that

> there is no compelling reason to restrict the attribution of action exclusively to humans and to social systems[101] (…). Personifying other non-humans is a social reality today and a political necessity for the future.

Writing from a systems theory perspective, Teubner sees attribution of personhood as a mechanism for social systems to reduce uncertainty: viewing a complex entity as a person enables you to communicate with it and to mutually establish expectations. In fact, 'through personification, the social system "parasitises" the intrinsic dynamics of autonomous processes in its environment.'[102] Rather than focusing on ontological properties (like mind, soul, reflexive capacities, empathy) as a condition for personhood, an entity is considered an 'actor' and attributed personhood by its environment under the minimum requirement of 'double contingency'. This means that in both directions of social interaction there is an element of unpredictability. We can treat non-humans as persons if there is 'a resistance, a "recalcitrance" which they [the non-humans] exert and which cannot be overcome by existing scientific knowledge.'[103] This seems to be the case with autonomic, adaptable software agents whose actions cannot be predicted in advance with sufficient precision by their owner or contracting party. Contrary to Latour who, in Teubner's account, seems to argue that all kinds of natural objects that 'modify a state of affairs by making a difference' can be treated as 'actants' in this way,[104] Teubner himself requires a 'capacity for dealing with proto-meaning'[105] as a precondition for personhood, and therefore considers adaptable software agents and domesticated animals as candidates for 'actants' with personhood.[106]

Besides 'actants', Teubner also embraces Latour's notion of 'hybrids', i.e. associations of human actors and non-human actants. Because some actants, like animals, lack certain communicative skills, they can team up with human actors – e.g. animal-rights groups – to function as full-blown actors in the social arena. These hybrids can now be personified as actors in their own right, under certain conditions. Although Teubner does not give the

---

[100] G. Teubner (2007), *Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law*, Max Weber Lecture Series 2007/04, European University Institute.
[101] I.e., legal persons like companies and states (author's footnote).
[102] Teubner, *supra* note 100 at 7.
[103] Teubner, *supra* note 100 at 12.
[104] For Latour's position, see *supra*, s. 1.3.1.
[105] Teubner does not explicitate what he means by 'dealing with proto-meaning'; it probably indicates a certain capacity, in a functional sense, to 'understand' communication with the environment, like a dog 'understands' a command to lie down without actually knowing human language, or like an adaptable software agent 'understands' that its owner is interested in English Victorian novels after consecutive commands to search for Austen, Eliot and Trollope.
[106] Teubner, *supra* note 100 at 12.

example, we can imagine electronic agents being employed by a company with limited liability as such a hybrid; rather than merely fitting the traditional model of a legal person (the company with limited liability), it is the hybrid of agents and company that should be the focus of our attention, because this empowers electronic agents to maximize their potential in the economic and social life.
The result of all this is that indeed non-humans gain access to social communication, albeit in a rather indirect way. The law plays a special role in this game; it stabilizes non-human personality by granting legal status to the hybrids via the construct of the juridical person, by attributing to them the capacity to act, by giving them rights, burdening them with duties and making them liable in several forms of legal responsibility.[107]

The attribution of legal status does not necessarily entail 'full-fledged legal subjectivity in order to open new political dynamics'.[108] Different gradations of legal personhood and legal capacity for action are possible, depending on the entity and the social context. In the case of animal rights, basically defensive institutions will be created for legal protection (to preserve ecology). In the case of electronic agents, however, legal personification, especially in economic and technological context, creates aggressive new action centers as basic productive institutions.'[109] In other words, attributing legal personhood, under certain conditions, to electronic agents capable of dealing with proto-meaning – adaptable, autonomic software agents –, or to hybrids of such agents and natural or legal persons, enables them to act in an economic and technologically significant way.

1.4.3.3 Matthias (2007)

Another 'plea for legal change' – as his subtitle emphasizes – is given by Andreas Matthias,[110] who has explored conditions for legal, moral, and social personhood and applied these to self-learning and self-adapting technology. He identifies an 'accountability gap' (*Verantwortungslücke*):

> there exists a growing class of accidents caused by machines, where the traditional ways of attributing responsibility are no longer compatible with our feeling of justice and the moral preconditions of society, since no-one has sufficient *control* over the actions of the machine, to be able to take responsibility.[111]

Autonomic agents – not only software agents, but also for example unmanned aerial vehicles or digital pets – present such an accountability gap, first, because they are unpredictable (*unberechenbar*) and, second, because they act outside the 'visibility horizon' of their maker, so that in case of failure, no manual intervention is possible.[112]

Matthias articulates five (cumulative) conditions for the ability to carry legal responsibility: intentionality, receptivity and responsiveness to causes, having second-order desires, legal sanity, and ability to distinguish between intended and merely foreseeable consequences of actions.[113] If these conditions are fulfilled, someone can carry legal responsibility, in the triple sense of having the capacity to perform legal acts (*Geschäftsfähigkeit*), to be held guilty of crimes (*Schuldfähigkeit*) and to be held accountable for unlawful acts (*Deliktfähigkeit*).[114]

---

[107] Teubner, *supra* note 100 at 16.
[108] Teubner, *supra* note 100 at 20.
[109] Teubner, *supra* note 100 at 20.
[110] A. Matthias (2007), *Automaten als Träger von Rechten. Plädoyer für eine Gesetzänderung*, diss. Berlin, Humboldt Universität, Berlin, Logos Verlag, 2007.
[111] Matthias, *supra* note 110 at 22 (emphasis in original, our translation).
[112] Matthias, *supra* note 110 at 37.
[113] Matthias, *supra* note 110 at 46 et seq.
[114] Matthias, *supra* note 110 at 63-72.

Interpreting the five conditions for legal responsibility in a functional way, he argues that legal accountability could accrue to certain classes of machines (software and/or hardware) – perhaps not current ones, but those in the foreseeable future that are even more self-learning and autonomic than today's machines. Indeed, it is 'only a matter of time before the distance between the producer/operator and the acting machine will be so large, that the absurdity will become obvious' of 'transferring the accountability to the producer or operator (who is ever less involved in the machine's acts)'.[115] This confirms Allen and Widdison's point that *not* attributing legal personhood to these machines is more of a legal fiction than providing for it would be.[116]

Matthias observes that 'persons' and 'human beings' should not be equated off-hand, since history and culture teach us that many humans were (and sometimes are) not considered by society or law as persons, and vice versa. Thus, Matthias' analysis warns us not to interpret criteria for personhood in an anthropomorphic way, but functionally in terms of whether the goals of legal accountability can be met. Machines can 'learn' and 'be educated' (e.g., through neural networks that can incorporate legal decisions into their rule system), and they can earn and administer money (since they perform economic tasks and can learn to manage bank accounts) out of which damages can be compensated. If autonomic machines are equipped with such tools that enable them to learn, administer money, and conclude insurance contracts, applying coercive powers (like education or compensating damages) to machines is not at all absurd, but often a natural extension of the original application area of these powers that can fulfill the same goals as the legislature originally intended.[117]

Matthias even goes as far as to argue that the criminal goals of special and general prevention as well as retribution could be reached – by 'punishing' the machine –, but here, his analysis is rather brief and less convincing than elsewhere.[118] Altogether, the 'plea for legal change' to hold machines legally accountable in order to stop the accountability gap is more convincing where it concerns the capacity to perform contracts and to be held accountable, to a certain extent, for tort. 'Geschäftsfähige' and 'deliktfähige' autonomic machines can in that sense well be attributed legal personhood.

1.5 Clarifying personhood and agency at different levels

1.5.1 Different types of personhood

It is helpful at this point to distinguish between various types of 'personhood'. Matthias makes a useful distinction between legal, moral, and social persons, with an increasing sense of 'personality'. That is, the widest class of persons is the legal person, i.e., a bearer of legal responsibility, like natural persons and juridical persons. They can contract and compensate for damages, and can also be the object of coercive or punitive measures, in a utilitarian – or functionalist – sense: they do not necessarily need to have a moral dimension. A narrower class is the moral person, i.e., those legal persons that are responsive to moral reasoning, like (most) human beings. They can be praised or detested, rewarded or punished, and they are

[115] Matthias, *supra* note 110 at 113-114 (our translation).
[116] Allen and Widdison, *supra* note 71.
[117] Matthias, *supra* note 110 at 239 et seq.
[118] Matthias, *supra* note 110 at 247-249. For special and general prevention, he does not explicitate how this would work with machines. As for retribution, Matthias argues (at 249, our translation) that, even if the machine does not observe retributive punishment as such, 'the only aspect important for the effectiveness of a retributive act is whether it makes the original victim experience an adequate feeling of satisfaction', and this could in principle also be effected by 'punishing' a machine. In our view, not only is this difficult to operationalize in practice, but focusing on the concrete victim's feelings also does not fit in criminal legal theory of retribution.

open to moral guilt. The narrowest class of persons, in Matthias' view, is the social person, also called the natural or 'full' person, i.e., the moral person who is socially accepted as a person. Most human beings are social persons, but not always; it is culturally dependent just which human beings are fully accepted in society as 'full' persons.[119]

We can extend Matthias' categorization with the model of virtual entities and abstract persons that has been developed in the FIDIS project.[120] An entity is 'anything that has a distinct existence; it is the fundamental "thing" that can be identified'; this includes physical entities (with some sort of physical constituency) and virtual entities, i.e., 'an entity which is or has been the product of the mind or imagination'. An abstract person is 'a virtual entity that can have rights, duties, obligations and/or responsibilities associated to it in a certain context'.[121] These rights or duties are not necessarily legal rights or duties – they can also be, for example, moral or technical in nature; if they are legal, however, then the abstract person is also a legal person in Matthias' sense. The legal person can therefore be seen as a subcategory of the category of abstract persons, which again is a subcategory of the category of virtual entities. This is illustrated in the following graph:

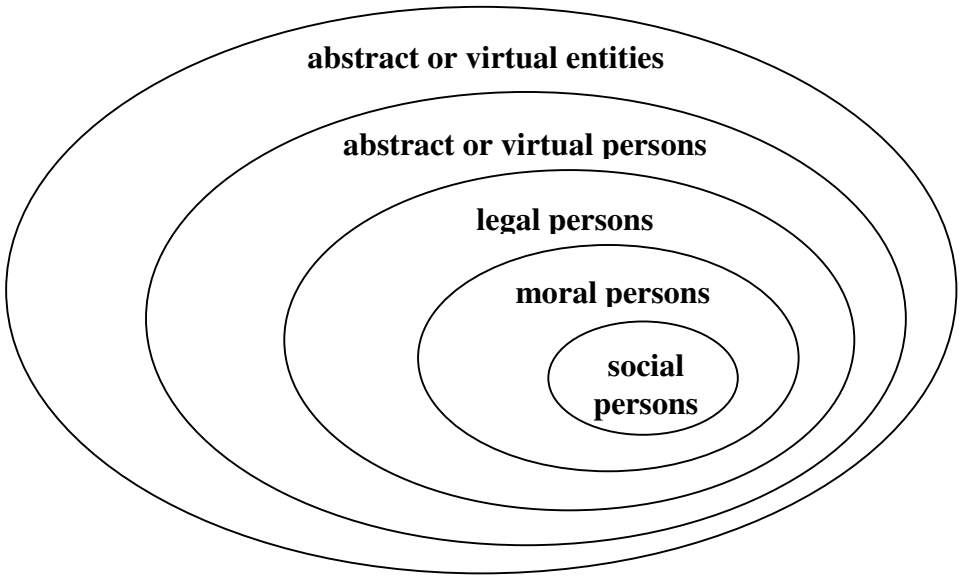

Figure 1. Categories of persons

One way of depicting the central issue of this article is illustrated with this graph: certain abstract entities, like pseudonyms, avatars, and software agents, operate sufficiently autonomically that they can be considered what some authors call abstract persons. The question we have explored would then be whether they could 'step up' one category and enter the more inner circle of legal persons, or perhaps even – in the long term – reach the category of moral or social persons.

Criteria for establishing legal personhood are not set in stone, and there is no obvious consensus distinguishable in legal literature what precisely is constitutive for legal personhood.[122] Some basics are clear, however, namely that personhood is associated with the legal capacity to act, and that this capacity involves civil actions (such as contracting) and

[119] Matthias, *supra* note 110 at 43-44.
[120] Jaquet-Chiffelle *supra* note 10.
[121] Jaquet-Chiffelle *supra* note 10 at 33-35.
[122] *Cf.* Matthias, *supra* note 110, p. 46, noting that many authors, while giving substantially varying criteria, each believe they have articulated the one and only sufficient condition for legal personhood (often based on an anthropomorphic paradigm of personhood).

criminal actions (committing a crime). For personhood to be meaningful, that means that an entity should be capable of performing such actions and bearing the consequences of them, which is particularly relevant when something goes wrong. It is here that legal personhood can be split in two:

- legal persons who are capable of civil actions, such as contracting, and who can bear consequences of civil wrong-doing: compensate for damages in case of breach of contract and tort; this may also include other unlawful but not morally wrong behavior, like misdemeanors[123] and administrative offences;
- legal persons who are capable of all types of legal actions, and who can bear both civil and criminal responsibilities; this is the category of legal persons who are also moral persons.

Thus, we can distinguish between a limited and a full sense of legal personhood. What is considered constitutive for these types of personhood may depend on one's perspective on the law, for example, whether one approaches the law from systems theory, functionalism, naturalism, or legal positivism.

1.5.2 Different types of agency

For the sake of clarification, we have introduced above conceptual distinctions, from a theoretical point of view, between different levels of automation and autonomy: automatic, autonomic and autonomous entities.[124] From a practical point of view, these distinctions might become less clear and difficult to assess in some specific situations. Indeed, we might observe some overlapping between these three types of automation and autonomy: for example, a software agent that is not able to modify its own program in order to achieve a certain goal is typically an automatic agent, but if its behavior appears to be unpredictable anyway, which can easily occur with complex automatic agents, it can function in much the same way as an autonomic agent. Probabilistic algorithms using an external source of entropy can lead to fixed algorithms with unpredictable behavior: the program's actions are not always predictable. This creates almost a continuum between automatic and autonomic entities. The decision to qualify a particular entity as automatic or autonomic might eventually depend on an arbitrary threshold.

The same is likely to happen with robots and cyborgs and other enhanced humans. The limit between human beings and machines might become fuzzy, creating a continuum too between autonomic and autonomous entities. When does a cyborg stop being part of mankind, and would it lose its autonomous property, when we replace, little by little, original parts of a human being by artificial components?

---

[123] Criminal offences consist of crimes and misdemeanours. Crimes are offences that harm some fundamental value and thus can be considered as morally wrong; misdemeanours are offences that breach a rule that is not primarily based on fundamental values but rather on creating order in society, such as the rule that cars drive on the right or left side of the road, or that citizens pay taxes.

[124] *Supra*, s. 1.3.2.

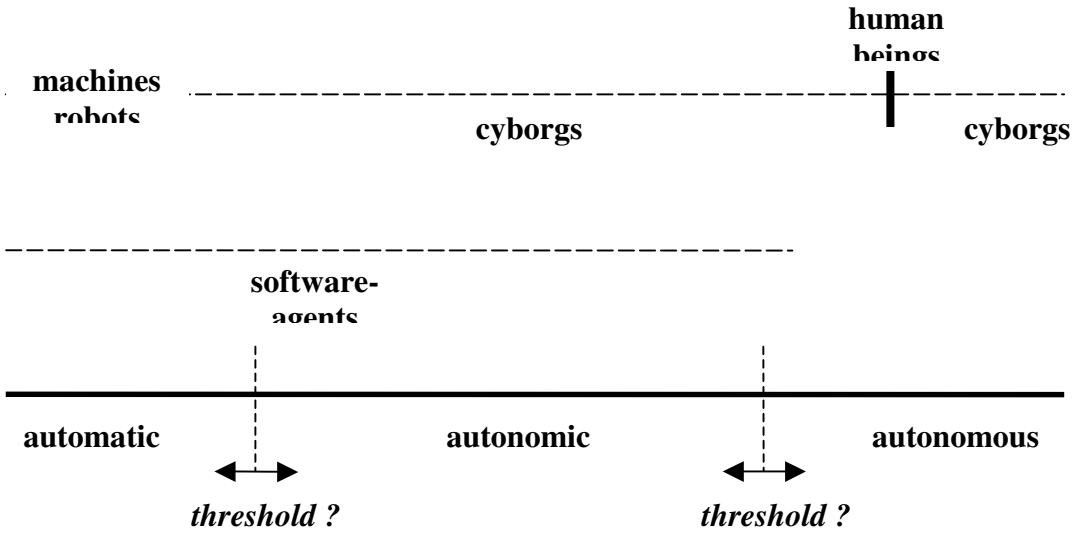

Figure 2. The continuum between automatic, autonomic and autonomous entities

Imagine the extreme case of a cyborg consisting of a living (enhanced) brain embodied in a robotical, artificial "body". What if moreover the control of this cyborg's actions is distributed between its human (enhanced) brain and either internal concurrent software agents or external computer programs? Would this affect the measure of freedom assumed in the attribution of culpable and wrongful action?

Above, in relation to the "natural person objection"[125] we have already indicated that the clear distinction between biological and non-biological entities may become blurred with the advent of cyborgs and synthetic biology. Even computers might move from pure electronic components to biological or hybrid ones. Research in biological computing is already a reality.[126] If we replace, step by step, artificial components of a computer or a robot by human parts, when does it become a person that could qualify as a human being?

The blurring of this distinction, however, does not imply that it becomes irrelevant. It rather shows that distinctions are analytical and usually do not map easily onto the flux of a fast changing reality. The famous question of when a collection of grains of sand counts as a heap comes to mind. In law we know that distinctions should not be made arbitrarily, since they will generate legal consequences. The point is to acknowledge that the difference between autonomic and autonomous action is neither simply given in reality nor an arbitrary social construction. This decision implies legal consequences that have ethical implications that need serious consideration if we wish to sustain fundamental (post)human rights that imply a measure of reflection and the capacity for intentional action.

The distinction between automatic and autonomic behavior on the one hand and autonomous action on the other hand, is particularly relevant for the question when a computer agent can be held liable for culpable and wrongful actions or can initiate an appeal to (post)human rights. As indicated above, the distinction is less – or maybe not at all – relevant for the question whether a restricted form of legal personhood can be created that allows the computer agent to act as a legal agent. It is to these questions that we now will try and provide an answer, building on insights from the literature review and the different types of personhood and agency that we have distinguished here.

---

[125] *Supra*, s. 1.4.1.3.
[126] See, for example, http://news.bbc.co.uk/2/hi/science/nature/358822.stm.

1.6 Meeting the challenge: computer agents as legal persons?

As we noted above,[127] what is considered constitutive for legal personhood depends on one's perspective on the law, for example, whether one looks at it from systems theory, functionalism, natural law, legal positivism, or from a relational conception of law.[128] Regardless of one's approach to the law, however, it is clear that emerging entities that operate at increasing distance from their principal pose a challenge to the law. This concerns first a challenge to *determine* the law, for instance, if an electronic agent, because of a software bug, buys a camera outside his supposed pre-programmed money range, is the contract null and void because of lack of intention to buy, or is it valid and should the principal pay – and can he then address the producer, programmer, or seller to compensate for his damages? Second, there is a challenge to *enforce* the law, because the distance between entity and principal – not only in the physical sense, but also in the metaphoric sense that the entity's action is not determined in detail by the principal's action – may make it hard to find the principal. Linking abstract persons' actions in the information society to their principals may require considerable effort, perhaps at a higher cost than the damage at issue. The third challenge concerns the point at which an autonomic agent develops a measure of autonomy that implies the capacity of intentional action, based on a measure of self-consciousness, even if this is hard to imagine today. Such autonomy would raise the question of whether such autonomous agents should have full standing in law, meaning that they can be called to account for criminal actions, while they are entitled to what we now call human rights. Facing this threefold challenge, the legal system has three potential courses of action, which can be seen as consecutive in time, although the different stages may, of course, overlap at certain points.

1.6.1 Short term: interpretation and extension of existing law

First, the actions of computer agents can be dealt with by interpreting and extending existing law, incorporating the new technical developments in the existing legal system. This is daily practice, and the law has an impressive tradition in construing ways to apply seemingly inappropriate provisions to seemingly new situations. To achieve the validity of contracts concluded by autonomic computers, the courts can qualify the general intention of the owner/user of the computer agent as sufficient for the intention that is required for individual contracts, creating the possibility for those who contracted with the computer agent to sue the 'principal' (note that since computer agents do not have legal personality at this point in time, the legal relation of principal and agent does not apply). This, however, will only work if the electronic agent is considered a tool in the hands of the owner/user, which – as we have seen above – is a legal fiction in as far as autonomic computers may decide to conclude contracts in ways the 'principal' cannot foresee with sufficient probability and which he has relatively little power to control by giving precise orders. Whereas viewing the autonomic agent as a tool could solve the problem of the contracting party, it may thus create substantial risks for the owner/user of this tool; and these risks could well hamper widespread adoption of overall useful autonomic computer agents. Additionally, we must acknowledge that the contracting party's problem will only be solved if – after tracing the 'principal' in the real world on the basis of the agent's data (which hopefully include correct identifying data of the principal) – it pays to file suit against this 'principal', who may in fact reside in another jurisdiction.

---

[127] *Supra*, s. 1.5.1.

[128] A relational theory of law emphasizes the instrumental as well as the constitutive aspects of law as two sides of the same coin. This is a normative position: the attribution of legal competence should always be instrumental as well as protective. This position means that legal personhood should be attributed in a manner that protects the relevant human or non-human entities. See *supra* note 31.

With today's computer agents, considering a computer agent as a tool nevertheless seems to work well enough for the time being. For those authors who claim that referring to contemporary computer agents as tools is a legal fiction,[129] other time-honored legal constructions within existing law may be preferable to address the risks and accountability problems. For example, rules can be – and in some jurisdictions have been – drafted for electronic agents,[130] stipulating under which conditions contracts are valid and who is liable for which actions of agents. Also, someone intending to use a computer agent and desiring to limit the risk of the agent acting unpredictably can establish a corporation to serve as the principal for the electronic agent.
For tomorrow's agents, however, applying and extending existing doctrines in these ways may stretch legal interpretation to the point of breaking, when Matthias' 'accountability gap' (see section 1.5.1) really emerges in practice.

1.6.2 Middle term: limited personhood with strict liability

Creative interpretation and novel sector-specific rules provide for legal certainty, and they can also – if the need to do so is felt – deviate from 'off-line' legal constructs, for example, limiting liability in order to stimulate the market for promising new technologies. However, at some point it may make more sense to introduce strict liability for electronic agents if their unpredictable actions are felt to be too risky for business or consumers.

In line with this strategy, interesting solutions have been suggested in the literature, notably to introduce a public register for agents, which could allow contractants to find the identity of an agent's principal, or, alternatively, to lay a claim on insurance for damages in case a registered agent goes haywire. The latter is similar to the establishment of victim funds, which is a way for society to control risks involving not too high losses for potentially many people, that are hard to attribute to individual causal actors.

A register of electronic agents might also be introduced together with a limited type of personhood for the electronic agents at issue. That is, the electronic agent itself will be responsible for its contracts and potential mishaps (outside of the moral or criminal sphere), based on strict liability. The agent could have money itself, for example by earning a small provision for each transaction it makes for his principal, and use this money – probably together with an insurance – to pay civil damages or administrative fines. It is currently not necessary to do this, but being aware of on-going technological developments that create more and more truly autonomic entities, we may have to consider this option in the middle term. Of course, this is not a trivial exercise; some kind of procedure will have to be developed for deciding which claims can be accepted, and which court has jurisdiction if the agent or the claimant does not agree. We also note Karnow's warning that polymorphous mobile electronic agents may be hard to identify. In that case liability of the agent itself does not solve the problem and may actually create a problem in the case that the principal is not liable because there is no way to locate the principal, or if the principal can claim that he never gave reason to believe that the transaction was concluded in its name.

Electronic agents are the most autonomic entities to date and thus the most likely candidate for 'stepping up' a category to become a legal person, under certain conditions. However, we should bear in mind that legal personhood has different functions: it allows an entity to function smoothly in social and economic interactions, and it provides it with legal protection. Different contexts may lead to different forms and scopes of legal personhood. Pseudonyms functioning as an entity in themselves, for example, will likely not become as autonomic as electronic agents, but they may acquire a 'personality' of their own (like Mark Twain, for

[129] *See supra* note 51 and surrounding text.
[130] *Cf.* UETA, UCITA and UECA, *supra* note 86.

example, is a better-known personality than his principal, Sam Clemens). The reputation gained by a pseudonym may make it economically attractive to allow trade of pseudonyms, or protection against defamation and slander in order to protect their commercial value. Although this can likely be effected very well with current laws and legal constructions, it could be worth exploring whether pseudonyms, if they indeed acquire an important societal function of their own, could not be given limited legal personhood, rather like a ship has been attributed legal personhood to solve the very complex interactions that ships have in global sea trade.

Also, perhaps a case could be made for comparing avatars to animals, and if the call for animal rights – often along with a plea for legal personhood for animals – continues to increase,[131] why could not avatars trigger a movement for avatar rights?[132] After all, people sometimes become very attached to their avatars,[133] and Tamagotchi and Paro[134] are examples of technological pets that appeal to people's emotions for their continued existence. Perhaps avatars and pet robots will become as cuddly as panda bears, and the social need to protect them from harm will lead legal scholars to argue for another type of limited legal personhood, in that they can defend themselves in court – at first represented by human beings, like companies are, but there seems no reason why, in principle, an avatar could not be represented by a lawyer-avatar. Echoing Teubner's provocative conclusion of his analysis of the ecological movement ('Trees do have standing'),[135] some scholar might, in twenty year's time, eloquently argue that, in the contemporary technological world, avatars or technological pets are so vital for social life that they need rights to legally protect them from harm, like ecological systems today, and hence 'avatar-human hybrids do have standing'. The problem, however, remains that since trees and animals are not capable of explaining their behavior in a court of law, granting them legal personhood seems a category mistake. Whereas granting restricted personhood could in fact make sense as a means of piecemeal engineering, this is not Teubner's piece of cake. Though we appreciate the provocative nature of his preference for sweeping statements, we think good reasons can be given against providing limited personhood for technological pets and criminal liability of animals[136].

1.6.3 Long term: full personhood with 'posthuman' rights

The constructions of limited legal personhood could evolve into the third strategy, namely to change the law more fundamentally by attributing full personhood to new types of entities. This would concern both liability on the basis of wrongful action and culpability and a lawful claim to posthuman rights. Can we imagine that computer agents should be attributed moral personhood in the long term, if they gain the ability to make moral (or moral-looking) decisions, based on self-consciousness (or something that looks to their environment like self-consciousness)? It seems important to distinguish between the issue of what standard is used to determine who or what qualifies for such full-fledged personhood and the issue of how we intend to establish whether a specific entity actually meets this criterion. We have argued that

---

[131] *Cf.* Teubner, *supra* note 100, with literature references.
[132] *Cf.* PETS, People for the Ethical Treatment of Software, http://www.elsop.com/wrc/humor/pets.htm, which parodies the animal-rights activist group PETA, People for the Ethical Treatment of Animals.
[133] *Supra*, note 8.
[134] *Supra*, s. 1.2.4.
[135] Teubner, *supra* note 100, p. 16. *Cf.* also Matthias' analysis of social personhood and machines: Matthias, *supra* note 110, p. 141-234.
[136] Note that times have existed when animals were considered eminently capable of criminal liability. Teubner, *supra* note 100, p. 1-2, starts his lecture with a lively report of the rats of Autun, who were tried before an ecclesiastical court for eating and wantonly destroying barley crops in the diocese. This took place in 1522. In modern legal systems, such trials seem rather less fitting.

the relevant criterion is the emergence of self-consciousness, since this allows us to address an entity as a responsible agent, forcing it to reflect on its actions as its own actions, which constitutes the precondition of intentional action. We note that intentional action – in this view – is an emergent property of an agent who is responding to the act of being called to account.[137] Evidently, autonomic agents also respond to their environment, but they do not respond by developing a reflection on their own action (in the German phrase: 'Reflektion auf eigenes Tun'), even if they may adapt their behavior to cope with changes in their environment. Some authors may object that this conclusion depends on the second issue, because the question is how we can establish that autonomic agents do not reflect on their behavior. Some cognitive scientists could even doubt whether what we call self-reflective, intentional, and autonomous action is not after all merely an epiphenomenon and an illusion of the brain.[138] Although this in itself entails an interesting discussion, it seems that *for law* the notion of an agent being capable of self-reflection and intentional action is crucial and does make a difference. For 'posthuman rights' to make sense, we have to assume that autonomous action exists, even if it exists only as a productive illusion.

We agree, then, with Solum that it makes no sense to exclude outright non-human entities from such rights and responsibilities. His point that such attribution should depend on the empirical finding that novel types of entities develop some kind of self-consciousness and become capable of intentional action seems reasonable, as long as we keep in mind that the emergence of such entities will probably require us to rethink notions of consciousness, self-consciousness and moral agency. Should a form of conscious self-reflection surface, then this will not necessarily be a property of a human-like robot (android), but rather erupt from distributed multi-agents systems that function as Karnow's polymorphous mobile agents. The intelligence and creativity of non-human entities presently depends on their interconnectedness, which allows for a measure of context awareness. In fact, cognitive science provides reasons to believe that human identity itself emerges from distributed brain processes, challenging the rationalist humanistic understanding of human agency. So, while non-human entities may at some point in future have a claim to 'posthuman' rights, our self-understanding may also evolve to seeing ourselves likewise as posthumans,[139] because we can no longer think of humans in the classic understanding of 'us' as a rational, unified identity that is transparent to itself.

1.7 Conclusion

To decide whether a specific entity qualifies as a person and the ensuing question of whether such artificial persons should qualify as legal abstract persons, we could take a relative approach. This means that next to establishing the preconditions for personhood we should acknowledge different levels of personhood, requiring different legal consequences. Thus, a particular smart application could qualify for a restricted form of legal personhood in as far as it can insure itself against liability; however, this should not imply the attribution of rights that

---

[137] *Cf.* Butler, J. (2005), *Giving an Account of Oneself*, New York: Fordham University Press; Duff, A., L. Farmer, S. Marshall and V. Tadros (2006), *The Trial on Trial. 2. Judgment and Calling to Account*, Oxford/ Portland, Oregon: Hart.

[138] *Cf. supra* note 35. This point also relates to the Turing test, see *supra* note 66.

[139] Hayles, *supra* note 46. Note that the loss of this rationalist subject does not, however, entail that we can no longer hold each other responsible, since responsibility does not hinge upon sovereignty of the self; see Butler (2005), *supra* note 137. On the relation between autonomic computing and human agency, see Hildebrandt, M., 'Autonomic and autonomous "thinking" as preconditions for criminal liability', in: Hildebrandt, M. and A. Rouvroy (eds.), *Autonomic Computing and Transformations of Human Agency. Philosophers of Law meeting Philosophers of Technology* (forthcoming). *Cf.* also the cyborg vision of self, *supra* note 59.

make no sense for an entity that has no consciousness, no intentionality, no feelings, no independent goals and no capacity for autonomous action. Criminal liability, which presumes a subject to be capable of autonomous action, had rather be attributed to another legal subject that does have this capability. Thus, while a non-human legal subject would be liable for harm caused in terms of private law, another legal subject would be liable for the same harm in terms of criminal law. This other legal subject could be a human being, a corporation or public body with legal personality.

What should interest us here is whether the attribution of a restricted legal personhood, involving certain civil rights and duties, has added value in comparison with other legal solutions. For several scholars, it makes sense to ponder future strategies to deal with new entities by attributing limited forms of personhood,[140] while others seem content with short-term interpretative solutions.[141] A few scholars even go further and argue that non-human entities, eventually, can lay a claim to full legal personhood.[142] Others, however, try to circumvent having to solve the legal problem of accountability by devising technical solutions that allow damages to be paid regardless of any determination of causation.[143] Choosing between these positions will depend largely on one's outlook on law and technology, on what constitutes a true 'person', and whether and to what extent the world is changing through the emergence of new types of entities.

For the time being, our research questions can be answered by the observation that interpretation and extension of the law seem to work well enough with today's computer agents. If technology evolves and entities like pseudonyms, avatars, and particularly electronic agents become more autonomic and acquire a 'personality' of their own, however, it might be useful to treat them as new entities with their own identities in themselves, with certain legal rights, duties, obligations, and/or responsibilities.[144] The majority view in the literature is that sooner or later, limited legal personhood with strict liability is a good solution for solving the accountability gap, particularly in contracting, and for electronic agents, this may be sooner rather than later. When it comes to attributing limited legal personhood involving rights to ensure legal protection, for example to protect pseudonyms or avatars, the literature is considerably more cautious; however, most literature to date tends to focus on electronic agents rather than on newer types of entities like pseudonyms, avatars, or pet robots, and perhaps the line of research on legal-protection rights for new types of entities has to be more fully developed in the literature.

When it comes to attributing full legal personhood and 'posthuman' rights to new types of entities, the literature seems to agree that this only makes sense if these entities develop self-consciousness. For the science-fiction-minded, it is interesting to speculate on a future where the independence of new entities, like androids or distributed multi-agent networks, reaches such a level that they move beyond autonomic-ness to a measure of autonomy, so that we may even consider giving them full legal personhood. But actually, this is a presumptuous statement. If networked machines begin to embody self-consciousness, considering their potential advantages over our own computing and acting capacities, it may well be *their*

---

[140] Solum, *supra* note 45; Allen and Widdison, *supra* note 71; Wettig and Zehendner, *supra* note 14.

[141] *Cf. supra*, s. 1.6.1.

[142] Matthias, *supra* note 110, and perhaps – the extent to which he advocates legal personhood is not entirely clear – Teubner, *supra* note 100.

[143] Karnow, *supra* note 88.

[144] *Cf.* Andrade et al., *supra* note 44 at 372, who conclude that ultimately, a 'choice must be made between the fiction of considering agents[’] acts as deriving from human's will and the endeavour of finding new ways of considering the electronic devices[’] own will and responsibility.'

decision whether or not to grant *us* legal personhood. Let us hope they will not treat us like we currently treat animals, as feed in our bio-industry.[145]

[145] This touches on debates around transhumanism; *Cf.* De Mul, *supra* note 58.