

# Ancilla Iuris

Digitale Rechtssubjekte? Zum privatrechtlichen  
Status autonomer Softwareagenten  
*Digital Personhood? The Status of Autonomous  
Software Agents in Private Law*

Gunther Teubner\*  
Translated by Jacob Watson

## Abstract

*Autonome Softwareagenten sind mathematisch formalisierte Informationsflüsse, denen schon heute in Wirtschaft und Gesellschaft unter bestimmten Bedingungen soziale Identität und Handlungsfähigkeit zugeschrieben wird. Aufgrund sozialer Handlungszurechnung sind sie zu nicht-menschlichen Mitgliedern der Gesellschaft geworden. Sie werfen drei neue Verantwortungsrisiken auf: (1) das Autonomierisiko, das in eigenständigen „Entscheidungen“ der Softwareagenten seinen Ursprung hat, (2) das Verbundrisiko, das auf die enge Kooperation von Menschen und Softwareagenten zurückzuführen ist, und (3) das Vernetzungsrisiko, das entsteht, wenn Computer in enger Verflechtung mit anderen Computern agieren. Diese Risiken stellen das Privatrecht vor die Herausforderung, für autonome digitale Informationssysteme einen neuen Rechtsstatus zu bestimmen – freilich nicht als volle Personifizierung von Softwareagenten, Mensch-Computer-Assoziationen oder Multi-Agenten-Systemen. Als Antwort auf die drei Risiken sollte vielmehr den Algorithmen jeweils ein auf ihre konkrete Rolle sorgfältig kalibrierter Rechtsstatus zuerkannt werden.*

*Drei neue Formen eines digitalen Rechtsstatus sind für autonome Softwareagenten angezeigt: (1) Akteur mit beschränkter Rechtssubjektivität, (2) Mitglied eines Mensch-Maschinen-Verbunds, (3) Teilelement eines Risikopools. (1) Für das Autonomierisiko ist es eine adäquate Antwort, den Softwareagenten den Status als teilrechtsfähige Akteure zuzuerkennen. Ihre autonomen Entscheidungen sind mit Rechtsverbindlichkeit auszustatten und sie sollten, falls rechtswidrig, Haftungsfolgen auslösen. Softwareagenten erhalten damit eine beschränkte Rechtssubjektivität, nämlich als Stellvertreter bindende Verträge für andere abzuschließen. Zugleich sind sie in Fällen vertraglicher und außervertraglicher Haftung als rechtlich handlungsfähige Hilfspersonen anzuerkennen, sodass das maschinelle Fehlverhalten selbst – und nicht bloß das Verhalten der dahinterstehenden Unternehmen – einen Pflichtverstoß darstellt, für den die Unternehmen einstehen müssen. (2) Eine denkbare Antwort auf das Verbundrisiko wäre ihr Rechtsstatus als Mitglied einer Mensch-Maschinen-Assoziation. De lege ferenda würde der Verbund selbst als rechtliches Zurechnungsobjekt für Handlungen, Rechte und Pflichten anerkannt werden. (3) Die Antwort auf das Vernetzungsrisiko schließlich wäre ein vom Haftungsrecht selbst autonom zu definierender Risikopool, der die Rechtsstellung der Algorithmen als Teil eines umfassenden digitalen Informationsflusses definieren und die Haftung des Pools ausschließlich am Ergebnis eines rechtswidrigen Verhaltens des Pools festmachen würde.*

## Abstract

*Autonomous software agents are mathematically formalized information flows. Already today in the economy and in society, they are attributed social identity and ability to act under certain conditions. Due to social action attribution, they have become non-human members of society. They pose three new risks: (1) the autonomy risk, which has its origin in stand-alone “decisions” taken by the software agents, (2) the association risk, which is due to the close cooperation between people and software agents, and (3) the network risk that occurs when computer systems operate in close integration with other computer systems. These risks pose a challenge for private law: to define a new legal status for autonomous digital information systems – however not simply as complete legal personification of software agents, human-computer associations or multi-agent systems respectively. Rather, in response to each of the three risks, a legal status should be granted to each of the algorithmic types that is carefully calibrated to their specific role.*

*Three new forms of digital legal status are presented for autonomous software agents here: (1) an actor with limited legal subjectivity, (2) a member of a human-machine association, (3) an element of a risk pool. (1) For autonomy, an adequate response would be to attribute limited legal personhood to software agents. (1) Their autonomous choices should be legally binding, and these should, if found to be illegal, trigger liability consequences. Software agents thereby are given a limited legal subjectivity, namely as representatives who may enter into contracts for others. At the same time, they are to be recognized as legally capable persons in cases of contractual and non-contractual liability so that the machine misbehavior itself – and not just the behavior of the underlying company – represents a breach of duty for which the company must stand. (2) A possible answer to the association risk would be their legal status as a member of a human-machine association. The association itself would be recognized de lege ferenda as the legal object of attribution for actions, rights and obligations. (3) Finally, the answer to the network risk would be that law itself would need to construct risk pools for the delimitation of these interrelationships. The risk pool would define the legal status of the algorithms as part of a comprehensive digital information flow, with the liability of the pool resulting only in the case of unlawful conduct of the pool.*

I.  
DIGITALISIERUNG:  
VERANTWORTUNGSLÜCKEN UND  
DOGMATIKDEFIZITE

1. Das Pfeifen im Walde oder: Die Angst vor dem  
Homo ex machina

„Verhaltensbiologisch dient das Pfeifen [...] als Geräuschmarkierung, wie es bei der Revierabgrenzung üblich ist, und soll Raubtiere, die dem Menschen gefährlich werden könnten, auf Distanz halten. Durch das Signalisieren des Revieranspruchs wird von eigener Unterlegenheitsangst abgelenkt.“ (Wikipedia) Die gefährlichen Raubtiere, die in die Reviere des Zivilrechts eindringen, sind nicht-kontrollierbare Algorithmen aus der digitalen Welt, Roboter, allgemeine Softwareagenten, die über hohe Intelligenz und Lernfähigkeit verfügen und für die Menschen neuartige ungeahnte Gefahren erzeugen. Doch mit selbstbewussten Geräuschmarkierungen halten viele juristische Autoren die digitalen Raubtiere auf Distanz: Sämtliche Gefährdungen durch die neuartige digitale Spezies, im Vertragsrecht wie im Haftungsrecht, könnten „mit tradierten rechtlichen Instrumenten bewältigt werden“.<sup>1</sup> Sie sehen keinerlei Anlass, von den bewährten Zurechnungskategorien, die das Computerverhalten stets als Menschenverhalten ausgeben, abzugehen. Softwareagenten könnten rechtlich ohne Probleme als bloße Maschinen, als willige Instrumente in den Händen ihrer menschlichen Herren behandelt werden.<sup>2</sup>

Stellt sich nicht schon heute mit Dringlichkeit das Problem, dass autonome Softwareagenten bisher nicht bekannte Risiken erzeugen, sodass ihnen nur mit neuen zivilrechtlichen Normen begegnet werden kann? Oder gar, dass den Softwareagenten ein eigenständiger privatrechtlicher Status zuerkannt werden muss? Letztlich entscheidend dürfte sein, ob die rasanten digitalen Entwicklungen schon jetzt schwer erträgliche Verantwortungslücken aufgerissen haben. Deshalb lautet die Gegenthese, die in der Literatur, aber auch vom Europäischen Parlament vertreten wird<sup>3</sup>: Unausweichlich verursachen Softwareagenten diese Verantwortungslü-

\* Gunther Teubner ist emeritierter Professor für Privatrecht und Rechtssoziologie an der Goethe-Universität Frankfurt. Für kritisch-konstruktive Kommentare dankt der Autor Marc Amstutz, Pasquale Femia, Malte Gruber und Dan Wielsch. Eine längere Fassung des Artikels erscheint in: *Archiv für civilistische Praxis* 218 (2018), 155–204.

1 Münchener Kommentar BGB/Busche, 7. Aufl. 2015, vor § 145 Rn. 37.

2 Bundesgerichtshof BGHZ 195, 126, Rn. 17; *Palandt/Ellenberger*, BGB 77. Aufl. 2018, vor § 116 Rn. 2; *Dieter Medicus/Jens Petersen*, Bürgerliches Recht, 26. Aufl. (2017), Rn. 256; *Christoph Sorge*, Softwareagenten: Vertragsschluss, Vertragsstrafe, Reugeld (2006), 36; *Stefan Kirn/Claus-Dieter Müller-Hengstenberg*, Technische und rechtliche Betrachtungen zur Autonomie kooperativ-intelligenter Softwareagenten, KI – Künstliche Intelligenz 29 (2015), 59–74, 67.

I.  
DIGITALIZATION:  
GAPS IN RESPONSIBILITY AND  
DEFICITS IN LEGAL DOCTRINE

1. The whistling in the woods or: The fear of the  
homo ex machina

“Behavioral biology describes whistling in the woods [...] as a sound marker, as it is customary when demarcating territory, and is intended to keep predators, that could become dangerous to humans, at a distance. By signaling the territorial claim one is distracted by one’s own fear of inferiority.” (Wikipedia) The dangerous predators that invade the territories of private law are uncontrollable algorithms from the digital world, robots, software agents with a high level of intelligence and the ability to learn which generate new kinds of undreamt-of dangers for humans. However, many legal scholars propose to keep digital predators at bay by making a lot of noise: All the hazards of these new digital species, within liability law as well as contract law, could according to them “be dealt with via traditional legal instruments”.<sup>1</sup> They see no reason to deviate from the established categories of attribution, which always imply that computer behavior is human behavior. Software agents could be legally treated without problems as mere machines, as willing instruments in the hands of their human masters.<sup>2</sup>

However, autonomous software agents are generating previously unknown risks, so that new rules in private law are probably needed to counteract them. Perhaps, software agents even have to be granted an independent status under private law. The decisive factor is likely to be whether the rapid digital developments have already opened up unacceptable gaps in responsibility. That is why the European Parliament and some legal authors advocate the counter-argument<sup>3</sup>: Software agents inevitably cause these gaps in responsibility, because their autonomous actions entail a massive loss of control for human actors. At the same time, how-

\* Gunther Teubner is Emeritus Professor of Law and Legal Sociology at the Goethe University Frankfurt. For critical and constructive comments the author would like to thank Marc Amstutz, Pasquale Femia, Malte Gruber and Dan Wielsch. An extended German version of this article appears in: *Archiv für civilistische Praxis* 218 (2018), 155–204.

1 For German legal doctrine, Münchener Kommentar BGB/Busche, 7th ed. 2015, BGB-AT, vor § 145 Rn. 37.

2 In Germany, Bundesgerichtshof BGHZ 195, 126, Rn. 17; *Palandt/Ellenberger*, BGB 77th ed. 2018, vor § 116 Rn. 2; *Dieter Medicus/Jens Petersen*, Bürgerliches Recht, 26th ed. (2017), Rn. 256; *Christoph Sorge*, Softwareagenten: Vertragsschluss, Vertragsstrafe, Reugeld (2006), 36; *Stefan Kirn/Claus-Dieter Müller-Hengstenberg*, Technische und rechtliche Betrachtungen zur Autonomie kooperativ-intelligenter Softwareagenten, KI – Künstliche Intelligenz 29 (2015), 59–74, 67.

cken, da ihr autonomes Handeln einen massiven Kontrollverlust menschlicher Akteure mit sich bringt. Zugleich aber nimmt die Abhängigkeit der Gesellschaft von autonomen Softwareagenten im großen Maßstab zu, und es ist äußerst unwahrscheinlich, dass die Gesellschaft auf deren Nutzung verzichtet.<sup>4</sup> Das Zivilrecht steht somit vor der Wahl: Entweder verleiht es autonomen Softwareagenten einen eigenständigen Rechtsstatus und behandelt sie selbst als verantwortliche Akteure oder es häufen sich immer mehr „Unfälle“, ohne dass ein Verantwortlicher dafür einsteht. Die Dynamik der Digitalisierung erzeugt unaufhörlich verantwortungsfreie Räume, die sich in Zukunft ausweiten werden.<sup>5</sup>

## 2. Konkrete Haftungslücken

Welches sind konkret die Verantwortungslücken? In der Informationswissenschaft werden folgende Sachverhalte genannt: Die Defizienzen entstehen in der Praxis dann, wenn die Software von Teams hergestellt werden, wenn Managemententscheidungen ebenso wichtig sind wie Programmierentscheidungen, wenn Dokumentationen über Anforderungen und Spezifizierungen eine große Rolle im resultierenden Code spielen, wenn trotz der Tests der Code-Genauigkeit viel von „*Off the shelf*“-Komponenten abhängt, deren Herkunft und Genauigkeit unklar ist, wenn die Performanz der Software das Resultat der begleitenden Kontrollen und nicht der Programmerstellung ist, wenn automatisierte Instrumente in der Konstruktion der Software benutzt werden, wenn die Arbeitsweise der Algorithmen von seinen *interfaces* oder sogar vom Systemverkehr beeinflusst wird, wenn die Software in nicht voraussehbarer Weise interagiert, wenn die Software mit Wahrscheinlichkeiten arbeitet oder über Anpassungsfähigkeit verfügt oder selbst das Resultat eines anderen Programmes ist.<sup>6</sup>

ever, society is becoming increasingly dependent on autonomous software agents on a large scale and it is highly unlikely that society will abandon their use.<sup>4</sup> Private law is therefore faced with a choice: Either it gives autonomous software agents an independent legal status and treats them as responsible actors, or there is an increasing number of “accidents”, without anyone being responsible for them. The dynamics of digitalization are constantly creating responsible-free spaces that will expand in the future.<sup>5</sup>

## 2. Specific liability gaps

What are the specific gaps in responsibility? In information science the following *scenarios* are mentioned: Deficiencies arise in practice when the software is produced by teams, when management decisions are just as important as programming decisions, when documentation of requirements and specifications plays a major role in the resulting code, when, despite the testing of code accuracy, a lot depends on “off-the-shelf” components whose origin and accuracy is unclear, when the performance of the software is the result of the accompanying controls and not of program creation, when automated instruments are used in the design of the software, if the operation of the algorithms is influenced by its *interfaces* or even by system traffic, if the software interacts in an unpredictable manner, if the software works with probabilities or has adaptability or is the result of another program.<sup>6</sup>

3 Europäisches Parlament (2016) „Entwurf eines Berichts mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik (2015/2103(INL))“, <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A8-2017-0005+0+DOC+PDF+V0//DE>, 5 ff. Auf das neuartige Autonomierisiko weisen hin: *Herbert Zech*, Zivilrechtliche Haftung für den Einsatz von Robotern: Zuweisung von Automatisierungs- und Autonomierisiken, in: Gless/Seelmann (Hg.), *Intelligente Agenten und das Recht*, 163–204 (2016), 175 f.; *Ruth Janal*, Die deliktische Haftung beim Einsatz von Robotern: Lehren aus der Haftung für Sachen und Gehilfen, in: Gless/Seelmann (Hg.), *Intelligente Agenten und das Recht* (2016), 139–162, 158; *Susanne Beck*, Der rechtliche Status autonomer Maschinen, *Aktuelle juristische Praxis* (2017), 183–191, 184.

4 *Andreas Matthias*, *Automaten als Träger von Rechten*. 2. Aufl. (2010), 15.

5 So die zentrale These von *Matthias* (Fn. 4), 111.

6 *Luciano Floridi/J.W. Sanders*, On the Morality of Artificial Agents, in: *Anderson/Anderson* (Hg.), *Machine Ethics* (2011), 184–212, 205.

3 European Parliament, <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0005+0+DOC+XML+V0//EN>. Pointing at the novel risk of autonomy: *Herbert Zech*, Zivilrechtliche Haftung für den Einsatz von Robotern: Zuweisung von Automatisierungs- und Autonomierisiken, in: Gless/Seelmann (ed.), *Intelligente Agenten und das Recht* (2016), 163–204, 175 f.; *Ruth Janal*, Die deliktische Haftung beim Einsatz von Robotern: Lehren aus der Haftung für Sachen und Gehilfen, in: Gless/Seelmann (ed.), *Intelligente Agenten und das Recht*, 139–162 (2016), 158; *Susanne Beck*, Der rechtliche Status autonomer Maschinen, *Aktuelle juristische Praxis* (2017), 183–191, 184.

4 *Andreas Matthias*, *Automaten als Träger von Rechten*. 2nd ed. (2010), 15.

5 This is *Matthias's* central thesis, (fn. 4), 111.

6 *Luciano Floridi/J.W. Sanders*, On the Morality of Artificial Agents, in: *Anderson/Anderson* (ed.), *Machine Ethics* (2011), 184–212, 205.

Im Recht hat das Auftauchen von Softwareagenten folgende Haftungslücken erzeugt:

*Computervernetzung:* Die am schwierigsten zu korrigierende Haftungslücke entsteht im Falle multipler Kausalität bei schädigenden Handlungen mehrerer Computer, wenn diese untereinander vernetzt sind. Mit den Haftungsnormen des geltenden Rechts kommt man hier nicht weiter.<sup>7</sup> Auch *de lege ferenda* ist noch kein brauchbarer Vorschlag ersichtlich. Im Falle des *High Frequency Trading* ist dieses Risiko offensichtlich geworden.<sup>8</sup> „Wer trägt das Risiko, wenn, wie am 6. Mai 2010, ein Flash-Crash ein gewaltiger Kursverlust des Dow Jones an der Börse in New York dadurch verursacht wurde, dass sich die Algorithmen der Wall Street, welche die Händlersysteme beherrschen, sich für einige Zeit unkontrolliert und unverständlich verhielten und einen Milliardenverlust bewirkten?“<sup>9</sup>

*Big Data:* Weitere Haftungslücken entstehen durch Fehleinschätzungen fehlerhafter Big Data-Vorhersagen. Big Data wird dazu eingesetzt, mithilfe außerordentlich großer bestehender Datenmengen zu prognostizieren, wie sich bereits bestehende gesellschaftliche Tendenzen oder Epidemien entwickeln und – notfalls – beeinflussen lassen.<sup>10</sup> Wenn der Grund für die fehlerhafte Berechnung nicht eindeutig festgestellt werden kann, bestehen die Schwierigkeiten, Kausalität und Verschulden zu bestimmen.

*Digitaler Vertragsbruch:* Wenn ein autonomer Softwareagent Vertragspflichten verletzt, dann soll, da er nicht über rechtliche Handlungsfähigkeit verfügt, nicht die Haftung für Erfüllungsgehilfen eingreifen. Stattdessen soll eine Haftung nur dann eintreten, wenn der menschliche Prinzipal selbst eine Vertragspflichtverletzung begeht. Hier entsteht eine problematische Haftungslücke: Kann der Betreiber nachweisen, dass der Softwareagent korrekt

In law, the emergence of software agents has created the following liability gaps:

*Computer networking:* The most difficult liability gap to correct arises in the case of multiple causality in the occurrence of damaging actions of several computers in a network. The liability norms of the applicable law do not go any further here.<sup>7</sup> There is also no sign of a useful proposal *de lege ferenda*. In the case of high frequency trading, this risk has become apparent.<sup>8</sup> “Who bears the risk if, as on May 6 2010, a flash crash caused a massive Dow Jones price drop on the New York Stock Exchange due to the algorithms of Wall Street, which control the trading systems, behaving for some time in an uncontrolled and incomprehensible manner and causing a loss of billions?”<sup>9</sup>

*Big Data:* Further liability gaps are caused by incorrect estimates of incorrect Big Data forecasts. Big Data is used to predict how existing societal trends or epidemics can develop and – if necessary – be influenced by extraordinarily large amounts of data.<sup>10</sup> If the reason for the faulty calculation cannot be clearly established, there are difficulties in determining causality and misconduct.

*Digital breach of contract:* If an autonomous software agent violates contractual obligations, the prevailing doctrine argues that the liability rules for auxiliary persons cannot be applied, as the agent does not have legal capacity to act. Instead, liability shall only arise if the human principal himself commits a breach of contract. This creates a problematic liability gap: If the operator can prove that the software agent has been used correctly without the operator him-

7 So deutlich *Indra Spiecker*, Zur Zukunft systemischer Digitalisierung: Erste Gedanken zur Haftungs- und Verantwortungszuschreibung bei informationstechnischen Systemen – Warum für die systemische Haftung ein neues Modell erforderlich ist, *Computer und Recht* (2016), 698–704.

8 Dazu *Malte-Christian Gruber*, On Flash Boys and Their Flashbacks: The Attribution of Legal Responsibility in Algorithmic Trading, in: *Jankowska/Kulawiak/Pawelczyk* (Hg.), *AI: Law, Philosophy & Geoinformatics* (2015), 88–102, 100.

9 *Stefan Kirm/Claus-Dieter Müller-Hengstenberg*, Intelligente (Software-) Agenten: Von der Automatisierung zur Autonomie? – Verselbstständigung technischer Systeme, *MultiMedia und Recht* (2014), 225–232, 227.

10 *Wolfgang Hackenberg*, in: *Hoeren/Sieber/Holzengel* (Hg.), *Handbuch Multimedia-Recht: Rechtsfragen des elektronischen Geschäftsverkehrs* 45. Auflage, 39. EL 2014, Teil 16.7 Rn. 58.

7 So clearly, *Indra Spiecker*, Zur Zukunft systemischer Digitalisierung: Erste Gedanken zur Haftungs- und Verantwortungszuschreibung bei informationstechnischen Systemen – Warum für die systemische Haftung ein neues Modell erforderlich ist, *Computer und Recht* (2016), 698–704.

8 See *Malte-Christian Gruber*, On Flash Boys and Their Flashbacks: The Attribution of Legal Responsibility in Algorithmic Trading, in: *Jankowska/Kulawiak/Pawelczyk* (ed.), *AI: Law, Philosophy & Geoinformatics* (2015), 88–102, 100.

9 *Stefan Kirm/Claus-Dieter Müller-Hengstenberg*, Intelligente (Software-) Agenten: Von der Automatisierung zur Autonomie? – Verselbstständigung technischer Systeme, *MultiMedia und Recht* (2014), 225–232, 227.

10 *Wolfgang Hackenberg*, in: *Hoeren/Sieber/Holzengel* (ed.), *Handbuch Multimedia-Recht: Rechtsfragen des elektronischen Geschäftsverkehrs* 45th ed., 39. EL 2014, Teil 16.7 Rn. 58.

eingesetzt wurde, ohne dass der Betreiber selbst eine Vertragspflicht verletzt hat, dann ist der Betreiber von jeglicher Haftung befreit.<sup>11</sup> Soll dann der Kunde den vom Computer der Gegenseite verursachten Schaden tragen?

*Außervertragliche Haftung:* Im Bereich außervertraglicher Haftung tritt ein ähnliches Problem auf, weil es bei verschuldensabhängiger Haftung nur auf die Pflichtverletzung des Betreibers, Herstellers oder Programmierers ankommen soll. Halten die beteiligten Menschen diese Pflichten ein, dann trifft sie keine Haftung.<sup>12</sup> Selbst wenn man kompensatorisch ihre Sorgfaltspflichten überspannt und die Grenzen der Zumutbarkeit überschreitet, wird die Haftungslücke nicht geschlossen.<sup>13</sup> Bei Schädigungen durch autonome Roboter führt dies, anders als bei Kraftfahrzeugschädigungen, bei denen stets Gefährdungshaftung eingreift, dazu, dass der Geschädigte ohne Schutz bleibt.

*Gefährdungshaftung:* Selbst die rechtspolitischen Vorschläge, die *de lege ferenda* vorschlagen, Schädigungen durch Softwareagenten mit Gefährdungshaftungsnormen zu begegnen, können erhebliche Haftungslücken nicht vermeiden. Die Leitprinzipien der Gefährdungshaftung können schwerlich als Vorbild dienen, da sie nicht auf das digitale Risiko zugeschnitten sind.

*Computererklärungen:* Eine fragwürdige Haftungslage ergibt sich auch im Recht der rechtsgeschäftlichen Computererklärungen, wenn der Softwareagent selbst wie ein „*falsus procurator*“ fälschlich einen Dritten als Vollmachtgeber angibt. In solchen Fällen soll das Risiko vollständig beim Geschäftsherrn des Softwareagenten liegen. Dies sehen viele Autoren als eine übermäßige Belastung an, die sich nicht rechtfertigen lässt, besonders hoch bei distribuerter Aktion oder *self-cloning*.<sup>14</sup>

self having violated a contractual obligation, the operator is exempted from any liability.<sup>11</sup> Should the customer then bear the damage caused by the other party's computer?

*Non-contractual liability:* A similar problem arises in the area of non-contractual liability because, in the case of fault-based liability, it is only the breach of duty by the operator, manufacturer or programmer that triggers liability. If the people involved comply with these obligations, then there is no liability.<sup>12</sup> The liability gap will not be closed, even if the courts overstretch duties of care for human actors.<sup>13</sup> If the damage is caused by autonomous robots, the injured party will be without protection.

*Liability for industrial hazards:* Even the legal policy proposals, which propose *de lege ferenda* to compensate for damages caused by software agents with *industrial hazard* liability standards, cannot avoid substantial liability gaps. The guiding principles of strict liability can hardly serve as a role model since they are not tailored to the digital risk.

*Computer declarations:* An unsatisfactory liability situation also arises if the software agent itself, like a “*falsus procurator*”, misrepresents a third party as the principal. In such cases, the risk should lie entirely with the principal of the software agent. Many authors see this as an excessive burden that cannot be justified, especially when it comes to distributed action or *self-cloning*.<sup>14</sup>

11 Z. B. Münchener Kommentar BGB/Grundmann, 7. Aufl. 2016, § 278 Rn. 46; *Kirn/Müller-Hengstenberg* (Fn. 9), 16.

12 Z. B. *Gerald Spindler*, Digitale Wirtschaft – analoges Recht: Braucht das BGB ein Update?, *JuristenZeitung* 71 (2016), 805–856, 816; ebenso *Peter Bräutigam/Thomas Klindt*, Industrie 4.0, das Internet der Dinge und das Recht, *Neue Juristische Wochenschrift* (2015), 1137–1142, 1140.

13 Die Tendenz zur Überspannung von Pflichten kritisieren *Malte-Christian Gruber*, Bioinformatonsrecht: Zur Persönlichkeitsentfaltung des Menschen in technisierter Verfassung (2015); *Janal* (Fn. 3), 155.

14 *Giovanni Sartor*, Agents in Cyberlaw, Proceedings of the Workshop on the Law of Electronic Agents (LEA 2002), [http://www.cirfid.unibo.it/\\*sartor/sartorpapers/gstartor2002\\_agenti\\_software.pdf](http://www.cirfid.unibo.it/*sartor/sartorpapers/gstartor2002_agenti_software.pdf) (2003).

11 E.g. Münchener Kommentar BGB/Grundmann, 7th ed. 2016, BGB § 278 Rn. 46; *Kirn/Müller-Hengstenberg* (fn. 9), 16.

12 E.g. *Gerald Spindler*, Digitale Wirtschaft – analoges Recht: Braucht das BGB ein Update?, *JuristenZeitung* 71 (2016), 805–856, 816; also *Peter Bräutigam/Thomas Klindt*, Industrie 4.0, das Internet der Dinge und das Recht, *Neue Juristische Wochenschrift* (2015), 1137–1142, 1140.

13 Criticizing the trend toward overloading of duties, *Malte-Christian Gruber*, Bioinformatonsrecht: Zur Persönlichkeitsentfaltung des Menschen in technisierter Verfassung (2015); *Janal* (fn. 3), 155.

14 *Giovanni Sartor*, Agents in Cyberlaw, Proceedings of the Workshop on the Law of Electronic Agents (LEA 2002), [http://www.cirfid.unibo.it/\\*sartor/sartorpapers/gstartor2002\\_agenti\\_software.pdf](http://www.cirfid.unibo.it/*sartor/sartorpapers/gstartor2002_agenti_software.pdf) (2003).

### 3. Dogmatische Fehlkonstruktionen

Die Verantwortungsdefizite entstehen, solange die Dogmatik darauf besteht, auf die neuen digitalen Realitäten ausschließlich mit dem hergebrachten begrifflichen Instrumentarium zu reagieren. Um dann aber dennoch mit den digitalen Entwicklungen einigermaßen Schritt zu halten, sieht sie sich gezwungen, mit fragwürdigen dogmatischen Hilfskonstruktionen den bisher unbekanntem Softwareagenten zu begegnen. Im Vertragsrecht hält sie unbeirrt daran fest, dass nur natürliche Personen in der Lage seien, rechtsverbindliche Erklärungen abzugeben. Deshalb muss sie mit problematischen Fiktionen arbeiten. Im Recht der vertraglichen und außervertraglichen Haftung muss sie, wenn Schädigungen auf einen Mensch-Computer-Verbund zurückzuführen sind, das schadensauslösende Ereignis ausschließlich am Handlungsanteil des Menschen festmachen und ist dann nicht mehr in der Lage, im Detail die Haftungsvoraussetzungen zu bestimmen. Und mit der Übernahme der Regeln zur Gefährdungshaftung geht sie einerseits zu weit, andererseits nicht weit genug, weil sie das digitale Risiko wie das bloße Kausalrisiko einer gefährlichen Sache behandelt. Gegenüber den Vernetzungen von Multi-Agenten-Systemen schließlich besteht allgemeine Ratlosigkeit.

Wenn in diesen Fällen das Recht auf den Einsatz von autonomen Softwareagenten nur mit herkömmlichen dogmatischen Mitteln reagiert und damit Verantwortungslücken aufbrechen lässt, wird der Schaden nicht etwa kollektiv über die Gesellschaft verteilt, sondern es gilt ein unbarmherziges *casum sentit dominus*. Hier nun setzt eine massive Kritik ein. Die Schadensfolgen dem Geschädigten aufzubürden, wird zu Recht als rechtspolitisch verfehlt und als fundamental ungerecht kritisiert. Sofern in solchen Situationen das Versagen autonomer Softwareagenten haftungsfrei bleibt, setzt dies falsche Anreize für Betreiber, Produzenten und Programmierer.<sup>15</sup> Und es sinkt die gesellschaftliche Bereitschaft, das im Prinzip vielversprechende Potential autonomer Softwareagenten voll zu nutzen. Vor allem aber widerspricht es dem Gerechtigkeitspostulat des notwendigen Zusammenhangs von Entscheidung und Verantwortung.<sup>16</sup>

### 3. Doctrinal misconceptions

The lack of responsibility arises as long as legal doctrine insists on responding to the new digital realities exclusively with traditional conceptual instruments. However, in order to keep pace with the digital developments at least to a certain extent, doctrine sees itself forced to react to the hitherto unknown software agents with questionable auxiliary constructions. In the field of contract law, legal doctrine firmly maintains that only natural persons are in a position to make legally binding declarations. Therefore, contract law has to work with problematic fictions. In the law of contractual and non-contractual liability, if damages are attributable to a human-computer network, it must attribute the damage-causing event exclusively to the action share of the human being and is then no longer able to determine in detail the liability prerequisites. And the rules on strict liability on the one hand go too far and on the other hand not far enough because they treat the digital risk like the mere causal risk of a dangerous matter. Finally, there is general perplexity regarding the networking of multi-agent systems.

If, in these cases, the law reacts to the use of autonomous software agents only with conventional conceptual means and thus leaves gaps in responsibility open, the damage is not distributed collectively across society, but rather a merciless *casum sentit dominus* applies. This is where there is massive criticism. To impose the consequences of the damage on the injured party is rightly criticized as wrong in legal policy terms and as fundamentally unfair. Provided that in such situations the failure of autonomous software agents remains free of liability, this creates false incentives for operators, producers and programmers.<sup>15</sup> And society's willingness to fully exploit the promising potential of autonomous software agents is diminishing. Above all, however, it contradicts the postulate of justice that maintains a necessary connection between decision and responsibility.<sup>16</sup>

15 Horst Eidenmüller, The Rise of Robots and the Law of Humans, SSRN Electronic Journal, <https://ssrn.com/abstract=2941001> (2017), 8.

16 Europäisches Parlament (Fn. 3), 7.

15 Horst Eidenmüller, The Rise of Robots and the Law of Humans, SSRN Electronic Journal, <https://ssrn.com/abstract=2941001> (2017), 8.

16 European Parliament (fn. 3), 7.

#### 4. Volle Rechtssubjektivität?

Volle Rechtsfähigkeit für Softwareagenten – dies ist die heute vieldiskutierte Antwort zahlreicher Juristen und Politiker auf diese Defizite.<sup>17</sup> Das Europäische Parlament hat im Januar 2017 auf der Grundlage des Delvaux-Berichtes beschlossen, „einen speziellen rechtlichen Status für Roboter zu schaffen, damit zumindest für die ausgeklügeltsten autonomen Roboter ein Status als elektronische Personen mit speziellen Rechten und Verpflichtungen festgelegt werden könnte, dazu gehört auch die Wiedergutmachung sämtlicher Schäden, die sie verursachen“. Wenn Roboter autonome Entscheidungen fällen, dann sollen sie als „elektronische Personen“, als juristische Personen im Vollsinn, anerkannt werden.<sup>18</sup>

In der Literatur wird zur Kompensation der oben genannten Defizite vorgeschlagen, dass *E-persons* die Fähigkeit haben sollen, als Juristische Vollpersonen Willenserklärungen im fremden, aber auch im eigenen Namen abzugeben,<sup>19</sup> dass sie Eigentum haben, über Geld verfügen, Bankkonten besitzen und Zugang zu Kredit haben sollen. Ja, dass sie für ihre Transaktionen Provisionen kassieren und dies selbstverdiente Geld dazu benutzen, um für Schädigungen oder Ordnungswidrigkeiten aufzukommen.<sup>20</sup> Im Haftungsrecht wird eine genuine Eigenhaftung der Softwareagenten gefordert. Entweder

#### 4. Full legal personhood?

Full legal personhood for software agents – this is the much-discussed answer of many lawyers and politicians to these shortcomings.<sup>17</sup> The European Parliament decided in January 2017, on the basis of the Delvaux report, to establish a special legal status for robots, so that at least for the most sophisticated autonomous robots a status as electronic persons with special rights and obligations could be established, including the redress of all the damage they cause. When robots make autonomous decisions, they should be recognized as “electronic persons”, as legal entities in the full sense of the word.<sup>18</sup>

It has been suggested in the literature that, to compensate for the above-mentioned shortcomings, *e-persons* should have the ability to make declarations of intent as full legal entities, both in their own name and in the name of others,<sup>19</sup> that they own property, have money, have bank accounts and should have access to credit. Indeed, that they collect commissions for their transactions and use this self-earned money to pay for damages.<sup>20</sup> Liability law requires a genuine self-liability of the software agents. Either they are to be allocated a fund for this purpose under property rights, which is alimented

17 Jan-Philip Günther, Roboter und rechtliche Verantwortung: Eine Untersuchung der Benutzer- und Herstellerhaftung (2016), 251 ff.; Evan J. Zimmerman, Machine Minds: Frontiers in Legal Personhood, SSRN Electronic Journal <https://ssrn.com/abstract=2563965> (2015); Susanne Beck, Über Sinn und Unsinn von Statusfragen: Zu Vor- und Nachteilen der Einführung einer elektronischen Person, in: Hilgendorf/Günther (Hg.), Robotik und Gesetzgebung (2013), 239–260, 255 ff.; Eric Hilgendorf, Können Roboter schuldhaft handeln? Zur Übertragbarkeit unseres normativen Grundvokabulars auf Maschinen, in: Beck (Hg.), Jenseits von Mensch und Maschine (2012), 119–132, 127 f.; Steffen Wettig, Vertragsschluss mittels elektronischer Agenten (2010); Steffen Wettig/Eberhard Zehendner, The Electronic Agent: A Legal Personality under German Law? Proceedings of the Law and Electronic Agents Workshop (2003), 97–112; Matthias (Fn. 4) 83 ff.; Emily M. Weitzenboeck, Electronic Agents and the Formation of Contracts, International Journal of Law and Information Technology 9 (2001), 204–234; Tom Allen/Robin Widdison, Can Computers Make Contracts?, Harvard Journal of Law & Technology 9 (1996), 25–52, 34 ff.; Lawrence B. Solum, Legal Personhood for Artificial Intelligences, North Carolina Law Review 70 (1992), 1231–1283. Zur verfassungsrechtlichen Diskussion Jens Kersten, Menschen und Maschinen: Rechtliche Konturen instrumenteller, symbiotischer und autonomer Konstellationen, JuristenZeitung 70 (2015), 1–8, 7f.; Albert Ingold, Grundrechtsschutz sozialer Emergenz: Eine Neukonfiguration juristischer Personalität in Art. 19 Abs. 3 GG angesichts webbasierter Kollektivitätsformen, Der Staat 53 (2014), 193–226. Strikt ablehnend dagegen Hans-Peter Bull, Sinn und Unsinn des Datenschutzes (2016), 120. Kritisch auch Melinda Lohmann, Ein europäisches Roboterrecht: überfällig oder überflüssig, Zeitschrift für Rechtspolitik (2017), 168–171; Schaub (Fn. 3), 345 f.; S.M. Solaiman, Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy, Artificial Intelligence and Law 25 (2017), 155–179; Gerald Spindler, Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz – Braucht das Recht neue Haftungskategorien? Eine kritische Analyse möglicher Haftungsgrundlagen für autonome Steuerungen, Computer und Recht 31 (2015), 766–776.

18 Europäisches Parlament (Fn. 3), 13.

19 Z. B. Wettig/Zehendner (Fn. 17), 97 ff.; Bert-Jaap Koops/Mireille Hildebrandt/David-Olivier Jaquet-Chiffelle, Bridging the Accountability Gap: Rights for New Entities in the Information Society?, Minnesota Journal of Law, Science & Technology 11 (2010), 497–558, 559 ff.

20 Koops et al. (Fn. 19), 554; C. Karnow, The Encrypted Self: Fleshing out the Rights of Electronic Personalities, John Marshall Journal of Computer Information Law 13 (1994), 1–16.

17 Jan-Philip Günther, Roboter und rechtliche Verantwortung: Eine Untersuchung der Benutzer- und Herstellerhaftung (2016), 251 ff.; Evan J. Zimmerman, Machine Minds: Frontiers in Legal Personhood, SSRN Electronic Journal <https://ssrn.com/abstract=2563965> (2015); Susanne Beck, Über Sinn und Unsinn von Statusfragen: Zu Vor- und Nachteilen der Einführung einer elektronischen Person, in: Hilgendorf/Günther (Hg.), Robotik und Gesetzgebung (2013), 239–260, 255 ff.; Eric Hilgendorf, Können Roboter schuldhaft handeln? Zur Übertragbarkeit unseres normativen Grundvokabulars auf Maschinen, in: Beck (ed.), Jenseits von Mensch und Maschine (2012), 119–132, 127 f.; Steffen Wettig, Vertragsschluss mittels elektronischer Agenten (2010); Evan J. Zimmerman, Machine Minds: Frontiers in Legal Personhood, SSRN Electronic Journal <https://ssrn.com/abstract=2563965> (2015); Steffen Wettig/Eberhard Zehendner, The Electronic Agent: A Legal Personality under German Law? Proceedings of the Law and Electronic Agents Workshop (2003), 97–112; Matthias (fn. 4) 83 ff.; Emily M. Weitzenboeck, Electronic Agents and the Formation of Contracts, International Journal of Law and Information Technology 9 (2001), 204–234; Tom Allen/Robin Widdison, Can Computers Make Contracts?, Harvard Journal of Law & Technology 9 (1996), 25–52, 34 ff.; Lawrence B. Solum, Legal Personhood for Artificial Intelligences, North Carolina Law Review 70 (1992), 1231–1283. On the constitutional discussion Jens Kersten, Menschen und Maschinen: Rechtliche Konturen instrumenteller, symbiotischer und autonomer Konstellationen, JuristenZeitung 70 (2015), 1–8, 7f.; Albert Ingold, Grundrechtsschutz sozialer Emergenz: Eine Neukonfiguration juristischer Personalität in Art. 19 Abs. 3 GG angesichts webbasierter Kollektivitätsformen, Der Staat 53 (2014), 193–226. Strictly opposed, Hans-Peter Bull, Sinn und Unsinn des Datenschutzes (2016), 120. Critical as well Melinda Lohmann, Ein europäisches Roboterrecht: überfällig oder überflüssig, Zeitschrift für Rechtspolitik (2017), 168–171; Schaub (Fn. 3), 345 f.; S.M. Solaiman, Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy, Artificial Intelligence and Law 25 (2017), 155–179; Gerald Spindler, Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz – Braucht das Recht neue Haftungskategorien? Eine kritische Analyse möglicher Haftungsgrundlagen für autonome Steuerungen, Computer und Recht 31 (2015), 766–776.

18 European Parliament (fn. 3), 18.

19 E.g., Wettig/Zehendner (fn. 17), 97 ff.; Bert-Jaap Koops/Mireille Hildebrandt/David-Olivier Jaquet-Chiffelle, Bridging the Accountability Gap: Rights for New Entities in the Information Society?, Minnesota Journal of Law, Science & Technology 11 (2010), 497–558, 559 ff.

20 Koops et al. (fn. 19), 554; C. Karnow, The Encrypted Self: Fleshing out the Rights of Electronic Personalities, John Marshall Journal of Computer Information Law 13 (1994), 1–16.



soll ihnen dafür ein Fonds eigentumsrechtlich zugeordnet werden, der durch Zahlungen der Beteiligten (Hersteller, Programmierer, Betreiber, Nutzer) aufgefüllt wird, oder eine Versicherung soll die Eigenschulden des Agenten abdecken.<sup>21</sup> Nicht nur sollen sie zu Trägern subjektiver Rechte werden, sondern sogar eigene Grundrechte geltend machen können, Rechte auf Entfaltung der Persönlichkeit, auf Nicht-Diskriminierung, auf wirtschaftliche Entfaltungsfreiheit und vor allem das Recht auf Meinungsfreiheit.<sup>22</sup>

Doch muss dem deutlich widersprochen werden: Forderungen auf volle digitale Rechtssubjektivität gehen an der heutigen Realität vorbei. Wie schon bei allen eingangs geschilderten Verantwortungslücken deutlich wird, geht es bis heute gar nicht um eigennütziges Handeln der Automaten, sondern stets um deren fremdnütziges Handeln für Menschen oder Organisationen, besonders Wirtschaftsunternehmen.<sup>23</sup> Ökonomisch gesprochen handelt es sich um eine *Principal-agent-Beziehung*, in der der Agent zwar abhängig ist, aber dennoch autonom handelt.<sup>24</sup> Softwareagenten sind digitale Sklaven, aber Sklaven mit übermenschlichen Fähigkeiten.<sup>25</sup> Und der Sklavenaufstand muss verhindert werden.

Volle Rechtssubjektivität wäre erst dann angebracht, wenn ihnen in Wirtschaft und Gesellschaft Eigentum an Ressourcen zugeordnet würde, mit denen sie eigene Profitinteressen verfolgen. Softwareagenten treten aber – jedenfalls bisher – gar nicht als eigeninteressierte Handlungseinheiten auf, sondern stets in Interaktion mit Menschen, für deren Interessenverfolgung sie eingesetzt werden. Gegenüber verbreiteten Vorstellungen isolierten Handelns der Computer und dessen unaufhörlicher Zunahme ist schon heute nach Aussage von Experten die Verflechtung von digitalen und humanen Handlungen sehr viel häufiger als das Handeln der Algorithmen in Einsamkeit und Freiheit.<sup>26</sup> In Zukunft wird mit zunehmendem Einsatz von künstlicher Intelligenz auch Zahl und Intensität ihrer Interaktionen mit Menschen zunehmen. Der

by payments from the parties involved (manufacturers, programmers, operators, users), or an insurance policy ought to cover the agent's own debts.<sup>21</sup> Not only is it argued that should they become bearers of rights in private law, but that they should also be able to assert their own constitutional rights, rights to personal development, non-discrimination, freedom of economic development and, above all, the right to freedom of expression.<sup>22</sup>

But this must be clearly contradicted: Demands for full digital personhood are ignoring today's reality. As is already clear from all the above-mentioned responsibility gaps, to this day it is not at all a question of the machines acting in their own interest but rather always in the interest of humans or organizations, especially commercial enterprises.<sup>23</sup> Economically speaking, it is a principal-agent relationship in which the agent is dependent but autonomous.<sup>24</sup> Software agents are digital slaves, but slaves with superhuman abilities.<sup>25</sup> And the slave revolt must be prevented.

Full legal subjectivity would only be appropriate if they were given ownership of resources in the economy and society with which they pursue their own profit interests. However, software agents – at least so far – do not act as self-interested action units at all, but always in interaction with people for whose pursuit of interests they are used. Compared to widespread ideas of computers acting in isolation and the activity's incessant increase, according to experts, the interweaving of digital and human actions is already much more frequent than the acting of algorithms in isolation.<sup>26</sup> In the future, the number and intensity of their interactions with humans will increase with the increasing use of artificial intelligence. Thus, the trend may not be towards isolated digital agents at all, but rather more towards human-computer associations, so

21 Hilgendorf (Fn. 17), 127 f.; Matthias (Fn. 4), 244; Sartor (Fn. 14).

22 Karnow (Fn. 20); Ingold (Fn. 17), 205 ff.; Kersten (Fn. 17), 2 ff., 8; Zimmerman (Fn. 17), 34 ff.

23 So Jan-Erik Schirmer, Rechtsfähige Roboter, *JuristenZeitung* (2016), 660–666, 665. Sollten in Zukunft Algorithmen eigennützig handeln, dann wird unter funktionalen Gesichtspunkten an eine Erweiterung ihrer Teilrechtsfähigkeit zu denken sein.

24 Locus classicus Michael Jensen/William H. Meckling, Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, *Journal of Financial Economics* 3 (1976), 306–360.

25 Kein Wunder, dass auf die Rechtsstellung der Sklaven im römischen Recht angesichts der parallelen Situation häufig verwiesen wird, dazu Jan D. Harke, Sklavhalterhaftung in Rom, in: Gless/Seelmann (Hg.), *Intelligente Agenten und das Recht* (2016), 97–118.

26 Argyro Karanasiou/Dimitris Pinotsis, Towards a Legal Definition of Machine Intelligence: The Argument for Artificial Personhood in the Age of Deep Learning, [https://www.researchgate.net/publication/316789688\\_Towards\\_a\\_Legal\\_Definition\\_of\\_Machine\\_Intelligence\\_The\\_Argument\\_for\\_Artificial\\_Personhood\\_in\\_the\\_Age\\_of\\_Deep\\_Learning](https://www.researchgate.net/publication/316789688_Towards_a_Legal_Definition_of_Machine_Intelligence_The_Argument_for_Artificial_Personhood_in_the_Age_of_Deep_Learning) (2017), 7f.

21 Hilgendorf (fn. 17), 127 f.; Matthias (fn. 4), 244; Sartor (fn. 14).

22 Karnow (fn. 20); Ingold (fn. 17), 205 ff.; Kersten (fn. 17), 2 ff., 8; Zimmerman (fn. 17), 34 ff.

23 So Jan-Erik Schirmer, Rechtsfähige Roboter, *JuristenZeitung* (2016), 660–666, 665. If algorithms are to act in their self-interest in the future, then an extension of their partial legal capacity will have to be considered from a functional point of view.

24 Locus classicus Michael Jensen/William H. Meckling, Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, *Journal of Financial Economics* 3 (1976), 306–360.

25 No wonder that the legal status of slaves in Roman law is often referred to in view of the parallel situation, Jan D. Harke, Sklavhalterhaftung in Rom, in: Gless/Seelmann (ed.), *Intelligente Agenten und das Recht* (2016), 97–118.

26 Argyro Karanasiou/Dimitris Pinotsis, Towards a Legal Definition of Machine Intelligence: The Argument for Artificial Personhood in the Age of Deep Learning, [https://www.researchgate.net/publication/316789688\\_Towards\\_a\\_Legal\\_Definition\\_of\\_Machine\\_Intelligence\\_The\\_Argument\\_for\\_Artificial\\_Personhood\\_in\\_the\\_Age\\_of\\_Deep\\_Learning](https://www.researchgate.net/publication/316789688_Towards_a_Legal_Definition_of_Machine_Intelligence_The_Argument_for_Artificial_Personhood_in_the_Age_of_Deep_Learning) (2017), 7f.

Trend geht also womöglich gar nicht in Richtung isolierter digitaler Agenten, sondern eher stärker in Richtung von Mensch-Computer-Verbänden, sodass immer weniger das Handeln der einzelnen Akteure selbst, sondern eher das Gesamthandeln des Hybrids oder ein umfassender Vernetzungszusammenhang der Gegenstand von juristischen Analysen sein muss.<sup>27</sup>

In solchen Mensch-Maschinen-Interaktionen ausschließlich den Automaten Rechte und Pflichten zuzurechnen, wie es den Vorschlägen zur vollen Rechtssubjektivität vorschwebt, wird folglich ebensowenig ihrer Rolle gerecht wie der Rolle der beteiligten Menschen. Es unterschlägt tendenziell den Beitrag der Menschen zum gesamten Handlungszusammenhang und verfehlt deren Vermögen als Haftungspotential. So wie bisher in Wirtschaft und Gesellschaft Softwareagenten eingesetzt werden, ist also weder ihre volle Rechtssubjektivität noch gar ihre Beförderung zu Juristischen Personen notwendig, stattdessen sind weniger voraussetzungsvolle Rechtskonstruktionen gefragt. Ihr Rechtsstatus müsste, wie es besonders gründlich Gruber herausgearbeitet hat, in einer funktionalen Sicht genau auf ihre dienende Rolle in Mensch-Maschinen-Interaktionen abgestimmt sein.<sup>28</sup> Autonome digitale Assistenz – für diese so präzisierte Rolle erscheint keine allgemeine Rechtsfähigkeit notwendig. Vielmehr stellt sich die Frage, ob – und wenn ja wie – dafür eine partielle Rechtssubjektivität von Softwareagenten anerkannt werden müsste.<sup>29</sup> Die heute die Diskussion beherrschende Alternative – entweder sind Softwareagenten bloße Instrumente oder sie sind vollwertige Rechtssubjekte – dürfte also falsch gestellt sein. Verfügt das Recht nicht über subtilere Konstruktionen, um den neuartigen digitalen Gefahren zu begegnen?

27 *Karanasiou/Pinotsi* (Fn. 26), 8. Zu den Schwierigkeiten, den Vernetzungsumfang rechtlich zu erfassen, *Malte-Christian Gruber*, Zumutung und Zumutbarkeit von Verantwortung in Mensch-Maschine-Assoziationen, in: Günther/Hilgendorf (Hg.), *Robotik und Gesetzgebung* (2013), 123–163, 126, 144 f.

28 *Malte-Christian Gruber*, Was spricht gegen Maschinenrechte?, in: Gruber/Bung/Ziemann (Hg.), *Autonome Automaten: Künstliche Körper und künstliche Agenten in der technisierten Gesellschaft* (2015), 191–206, 198; *Malte-Christian Gruber*, Rechtssubjekte und Teilrechtssubjekte des elektronischen Geschäftsverkehrs, in: Beck (Hg.), *Jenseits von Mensch und Maschine: Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs* (2012), 133–160. Ebenso dezidiert *Vanessa Kluge/Anne-Kathrin Müller*, *Autonome Systeme: Überlegungen zur Forderung nach einer ‚Roboterhaftung‘*, *InTeR Zeitschrift zum Innovations- und Technikrecht* 5 (2017), 24–31, 27; *Schirmer* (Fn. 23), 660, 663ff.; *David Gindis*, *Legal Personhood and the Firm: Avoiding Anthropomorphism and Equivocation*, *Journal of Institutional Economics* 12 (2016), 499–513, 17.

29 Beschränkte Rechtsfähigkeit elektronischer Agenten in Bezug auf Stellvertretung und auf Gehilfenhaftung wurde vorgeschlagen in *Gunther Teubner*, *Elektronische Agenten und große Menschenaffen: Zur Ausweitung des Akteursstatus in Recht und Politik*, *Zeitschrift für Rechtssoziologie* 27 (2006), 5–30. Der Vorschlag soll im Folgenden im Detail ausgearbeitet werden, besonders unter III, 1–4. Für die beschränkte Rechtsfähigkeit von Softwareagenten plädieren auch *Gruber* (Fn. 13), 227 ff.; *Schirmer* (Fn. 23), 663 ff.; *Koops et al.* (Fn. 19), 512 f., 559; implizit *Kefler* (Fn. 95), 592 f.; *Paulius Čerkaa/Jurgita Grigienė/Gintarė Sirbikytė*, *Is it Possible to Grant Legal Personality to Artificial Intelligence Software Systems?*, *Computer Law & Security Review* 33 (2017), 685–699.

that it is less the individual actors' actions themselves and instead the overall actions of the hybrid or a comprehensive computer network context that must be the subject of legal analysis.<sup>27</sup>

In such human-machine interactions, therefore, it is neither fair to assign rights and obligations exclusively to machines, as envisaged in the proposals for full legal subjectivity, nor does it do justice to their role, nor to the role of the people involved. It tends to undermine the contribution of humans to the whole context of action and misses their liability potential. In the same way that software agents have been used in business and society up to now, neither their full legal subjectivity nor their promotion to legal entities is necessary; instead, less ambitious legal constructions are required. Their legal status should, as Gruber has elaborated thoroughly, be precisely attuned to their role in human-machine interactions from a functional point of view.<sup>28</sup> Autonomous digital assistance – no general legal subjectivity appears to be necessary for this more precise role. Rather, the question arises as to whether – and if so, how – a partial legal subjectivity of software agents would have to be recognized.<sup>29</sup> The alternative that dominates today's debate – either software agents are mere instruments or they are fully-fledged legal entities – is therefore just wrong. Doesn't the law have more subtle constructions to counter the new digital threats?

27 *Karanasiou/Pinotsi* (fn. 26), 8. On the difficulties of understanding the networking aspect legally, *Malte-Christian Gruber*, *Zumutung und Zumutbarkeit von Verantwortung in Mensch-Maschine-Assoziationen*, in: Günther/Hilgendorf (ed.), *Robotik und Gesetzgebung* (2013), 123–163, 126, 144 f.

28 *Malte-Christian Gruber*, *Was spricht gegen Maschinenrechte?*, in: Gruber/Bung/Ziemann (ed.), *Autonome Automaten: Künstliche Körper und künstliche Agenten in der technisierten Gesellschaft* (2015), 191–206, 198; *Malte-Christian Gruber*, *Rechtssubjekte und Teilrechtssubjekte des elektronischen Geschäftsverkehrs*, in: Beck (ed.), *Jenseits von Mensch und Maschine: Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs* (2012), 133–160. Also decidedly *Vanessa Kluge/Anne-Kathrin Müller*, *Autonome Systeme: Überlegungen zur Forderung nach einer ‚Roboterhaftung‘*, *InTeR Zeitschrift zum Innovations- und Technikrecht* 5 (2017), 24–31, 27; *Schirmer* (fn. 23), 660, 663ff.; *David Gindis*, *Legal Personhood and the Firm: Avoiding Anthropomorphism and Equivocation*, *Journal of Institutional Economics* 12 (2016), 499–513, 17.

29 Limited legal personhood of electronic agents with regard to representation and auxiliary liability has been proposed by *Gunther Teubner*, *Elektronische Agenten und große Menschenaffen: Zur Ausweitung des Akteursstatus in Recht und Politik*, *Zeitschrift für Rechtssoziologie* 27 (2006), 5–30. The proposal will be worked out in detail below, in particular under III, 1–4. Also pleading for the limited legal capacity of software agents, *Gruber* (fn. 13), 227 ff.; *Schirmer* (fn. 23), 663 ff.; *Koops et al.* (fn. 19), 512 f., 559; implizit *Kefler* (fn. 95), 592 f.; *Paulius Čerkaa/Jurgita Grigienė/Gintarė Sirbikytė*, *Is it Possible to Grant Legal Personality to Artificial Intelligence Software Systems?*, *Computer Law & Security Review* 33 (2017), 685–699.

## II. AKTANTEN: AUTONOMIERISIKO

Die Frage, ob und wie den Softwareagenten der neuartige Rechtsstatus einer strikt funktional definierten Rechtssubjektivität zuerkannt werden soll, kann von der bloßen Aufdeckung der eingangs geschilderten Verantwortungslücken allein nicht geklärt werden. Denn diese sind nur die Symptome von tieferliegenden Risiken, die vom Einsatz autonomer Softwareagenten ausgelöst werden und auf die das Recht mit einem neuartigen Rechtsstatus reagieren sollte.<sup>30</sup> Drei Risiken der Digitalität stehen im Vordergrund: (1) das Autonomierisiko, das aus eigenständigen „Entscheidungen“ der Softwareagenten entsteht, (2) das Verbundrisiko, das auf die enge Kooperation von Mensch und Softwareagent zurückzuführen ist, und (3) das Vernetzungsrisiko, das entsteht, wenn Computer nicht isoliert agieren, sondern in enger Verflechtung mit anderen Computern. Erst deren genauere Analyse kann Richtungsangaben dafür machen, welcher privatrechtliche Status Softwareagenten zuerkannt werden sollte. Im Folgenden soll eingehend das Autonomierisiko und dessen rechtliche Konsequenzen diskutiert werden, während die anderen beiden Risiken nur eher cursorisch behandelt werden können.

Das Autonomierisiko, das vom prinzipiell unvorhersehbaren Verhalten selbstlernender Algorithmen erzeugt wird, verlangt nach anderen Formen der Risikoabsorption als das schon länger bekannte bloße Automatisierungsrisiko.<sup>31</sup> Es verwirklicht sich dann, wenn Softwareagenten tatsächlich, wie in den beteiligten Wissenschaften, besonders der Informationsphilosophie, der Aktor-Netzwerk-Theorie und der Systemtheorie, vielfach diskutiert wird, als handlungsfähige Akteure auftreten.

### 1. Anthropomorphisierung?

Bedeutet dies, wie oft behauptet, dass Computer kurzschlüssig mit Menschen gleichgesetzt werden? Um nicht falschen Anthropomorphisierungen aufzusitzen, muss man die digitale Handlungsfähigkeit erst in ihrer Eigenart verstehen. Dazu ist es notwendig, die Parallele zur Handlungsfähigkeit anderer nicht-menschlicher Akteure, zu den als juristischen Personen verfassten formalen Organisationen, zu ziehen. Um das Ergebnis einer weitverzweigten Diskussion zum Substrat der juristi-

<sup>30</sup> Zech (Fn. 3), 172 ff.

<sup>31</sup> Zech (Fn. 3), 175 f.; Janal (Fn. 3), 158.

## II. ACTANTS: AUTONOMY RISK

The question of whether and how the software agents should be granted the new legal status of a strictly functionally defined legal subjectivity cannot be clarified by the mere discovery of the above-mentioned gaps in responsibility alone. These are only the symptoms of underlying risks that are triggered by the use of autonomous software agents and to which the law should respond with a new legal status.<sup>30</sup> Three risks of digitality are at the forefront: (1) the autonomy risk that arises from independent “decisions” of the software agents, (2) the network risk that is due to the close cooperation between humans and software agents, and (3) the network risk that arises when computers do not act in isolation but in close interdependence with other computers. Only a more detailed analysis of these can give directions as to what private law status should be granted to software agents. In the following, the risk of autonomy and its legal consequences will be discussed in detail, while the other two risks can only be treated in a more cursory manner.

The autonomy risk generated by the principally unpredictable behavior of self-learning algorithms calls for other forms of risk absorption than the automation risk that has been known for some time.<sup>31</sup> It becomes a reality when software agents actually dispose of action capacity, as is often discussed in the sciences involved, especially information philosophy, actor-network theory and systems theory.

### 1. Anthropomorphism?

Does this mean, as often claimed, that computers are equated with people? In order not to mount false anthropomorphisms, one has to understand the peculiarity of the digital capacity for action. To this end, it is necessary to draw parallels with the capacity of other non-human actors to act, with the formal organizations drawn up as legal entities. To briefly mark the outcome of a wide-ranging discussion on the legal entity’s substrate: We must move away from the familiar idea that the social substrate

<sup>30</sup> Zech (fn. 3), 172 ff.

<sup>31</sup> Zech (fn. 3), 175 f.; Janal (fn. 3), 158.

schen Person kurz zu kennzeichnen: Man muss sich von der vertrauten Vorstellung lösen, das soziale Substrat juristischer Personen sei eine Vielheit konkreter Menschen.<sup>32</sup> Als Substrat kommen nicht die üblichen Verdächtigen in Betracht, weder Otto von Gierkes berüchtigte „reale Verbandspersönlichkeit“, noch Emile Durkheims „Kollektivbewusstsein“, noch James Colemans „Ressourcenpool“ und auch nicht Maurice Haurious „Institutionen“.<sup>33</sup> Der Kollektivakteur, wie ihn Talcott Parsons, Niklas Luhmann und andere definieren, ist nicht eine Gruppe von Individuen, sondern eine Kette von Mitteilungen. Organisationen sind nicht Häuser oder Menschengruppen oder Ressourcenpools, sondern Entscheidungsketten. Unter der doppelten Voraussetzung, dass erstens eine solche Kommunikationskette über sich selbst kommuniziert, also eine Selbstbeschreibung herstellt, und dass zweitens gerade dieser Selbstbeschreibung kommunikative Ereignisse als von ihr selbst vollzogene Handlungen zugeschrieben werden, entsteht die soziale Realität eines Kollektivakteurs, das soziale Substrat der juristischen Person.<sup>34</sup>

In ganz paralleler Weise sind auch Softwareagenten als Algorithmen, also als mathematisch formalisierte Informationsflüsse zu verstehen, denen in Wirtschaft und Gesellschaft unter bestimmten Bedingungen soziale Identität und Handlungsfähigkeit zugeschrieben wird.<sup>35</sup> Die Parallele wird deutlich, wenn man zwei Fehlvorstellungen der Personifizierung nicht-menschlicher Entitäten zurückweist: Falsch ist die Vorstellung, dass im Falle von Organisationen eine Menschenansammlung sich zu einer organisatorischen realen

of legal persons is a multiplicity of concrete people.<sup>32</sup> The substrate is not one of the usual suspects, neither Otto von Gierke’s notorious “real collective personality”, nor Emile Durkheim’s “collective consciousness”, nor James Coleman’s “resource pool”, nor Maurice Haurious’s “institution”.<sup>33</sup> The collective actor as defined by Talcott Parsons, Niklas Luhmann and others is not a group of individuals – but a chain of messages. Organizations are not buildings or groups of people or resource pools, but decision chains. The social reality of a collective actor, the social substrate of the legal person, arises from the dual premise that first such a communication chain communicates about itself, i.e. creates a self-description, and second that it is this self-description that communicative events are ascribed to it as actions.<sup>34</sup>

In a completely parallel manner, software agents are to be understood as algorithms, i.e. as mathematically formalized information flows, to which social identity and the ability to act are ascribed in the economy and society under certain conditions.<sup>35</sup> The parallel becomes clear when one rejects two misconceptions of the personification of non-human entities: It is wrong to imagine that, in the case of organizations, an association of people is aggregated into an organizationally real collective person. And just as wrong is the idea that in

32 Zum soziologischen Begriff des Kollektivakteurs *Niklas Luhmann*, *Soziale Systeme: Grundriß einer allgemeinen Theorie* (1984), 270 ff.; vgl. auch *Niklas Luhmann*, *Die Gesellschaft der Gesellschaft* (1997), 834 f.; *Niklas Luhmann*, *Die Politik der Gesellschaft* (2000) 241. Zum Verhältnis von Kollektivakteur und Juristischer Person untersucht *Gunther Teubner*, *Unternehmenscorporatismus: New Industrial Policy und das Wesen der Juristischen Person*, *Kritische Vierteljahresschrift für Gesetzgebung und Rechtswissenschaft* 2 (1987), 61–85. Eine detaillierte Analyse der Personifizierung nicht-menschlicher Entitäten bietet *Guido Sprenger*, *Communicated into Being: Systems Theory and the Shifting of Ontological Status*, *Anthropological Theory* 17 (2017), 108–132, 111. Zu durchaus ähnlichen Ergebnissen kommt die Diskursanalyse unter dem Stichwort der Subjektivierung, dazu *Thomas Alkemeyer/Ulrich Bröckling/Tobias Peter* (Hg.), *Jenseits der Person: Zur Subjektivierung von Kollektiven* (2018); *Tomas Marttila*, *Post-Foundational Discourse Analysis: A Suggestion for a Research Program*, *Forum: Qualitative Social Research* <http://www.qualitative-research.net/> 16 (2015), 4.3.

33 Zur „realen Verbandspersönlichkeit“ *Otto von Gierke*, *Das Wesen der menschlichen Verbände* (1902); zum „Kollektivbewusstsein“ *Emile Durkheim*, *Über die Teilung der sozialen Arbeit* (2004 [1883]), 121 ff.; zum „Ressourcenpool“ *James S. Coleman*, *Foundations of Social Theory* (1990), 325 ff.; zu „Normkomplexen“ *Maurice Hauriou*, *Die Theorie der Institution* (1965).

34 Dazu eingehend *Teubner* (Fn. 32).

35 Zur Personifizierung von Algorithmen in informationstheoretischer Sicht *Floridi/Sanders* (Fn. 6), 187 ff.; *Sabine Thürmel*, *The Participatory Turn: A Multidimensional Gradual Agency Concept for Human and Non-human Actors*, in: *Misselhorn* (Hg.), *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation* (2015), 45–62, 52 ff.; in öffentlich-rechtlicher Sicht *Ingold* (Fn. 17), 220 ff.; in privatrechtlicher Sicht *Teubner* (Fn. 29).

32 On the sociological concept of the collective actor, *Niklas Luhmann*, *Social Systems* (1995), 198 ff.; *Niklas Luhmann* (2012/2013) *Theory of Society*, Ch. 4, XIV; *Niklas Luhmann*, *Die Politik der Gesellschaft* (2000) 241. On the relationship between the collective actor and the legal entity, *Gunther Teubner*, *Enterprise Corporatism: New Industrial Policy and the 'Essence' of the Legal Person*, *The American Journal of Comparative Law* 36 (1988), 130–155. A detailed analysis of the personification of non-human entities offers *Guido Sprenger*, *Communicated into Being: Systems Theory and the Shifting of Ontological Status*, *Anthropological Theory* 17 (2017), 108–132, 111. The discourse analysis under the keyword subjectivization comes to very similar results *Thomas Alkemeyer/Ulrich Bröckling/Tobias Peter* (ed.), *Jenseits der Person: Zur Subjektivierung von Kollektiven* (2018); *Tomas Marttila*, *Post-Foundational Discourse Analysis: A Suggestion for a Research Program*, *Forum: Qualitative Social Research* <http://www.qualitative-research.net/> 16 (2015), 4.3.

33 On the conception of the real collective person, *Otto von Gierke*, *Das Wesen der menschlichen Verbände* (1902); on collective consciousness, *Emile Durkheim*, *The Division of Labor in Society* (1933 [1883]), 79 ff.; on resource pool, *James S. Coleman*, *Foundations of Social Theory* (1990), 325 ff.; on norm complexes, *Maurice Hauriou*, *Aux sources du droit: Le pouvoir, l'ordre et la liberté* (1933).

34 In detail, *Teubner* (fn. 32).

35 On the personification of algorithms from an information theory perspective, *Floridi/Sanders* (fn. 6), 187 ff.; *Sabine Thürmel*, *The Participatory Turn: A Multidimensional Gradual Agency Concept for Human and Non-human Actors*, in: *Misselhorn* (ed.), *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation* (2015), 45–62, 52 ff.; from a public law perspective, *Ingold* (fn. 17), 220 ff.; from a private law perspective, *Teubner* (fn. 29).

Verbandspersönlichkeit verdichtet. Und ebenso falsch ist die Vorstellung, dass im Falle von Softwareagenten ein Computer in einen *homo ex machina* transformiert wird. In beiden Fällen geht es exakt um das Gleiche: um die gesellschaftliche Zuschreibung von eigener Handlungsfähigkeit an Kommunikationsprozesse.

Den prinzipiellen Unterschied zur menschlichen Handlungsfähigkeit hat Bruno Latour in seiner berühmten Aktor-Netzwerk-Theorie, in der handlungsfähige digitale Prozesse mit dem Neologismus „Aktanten“ belegt werden, herausgearbeitet und zugleich die genaueren Bedingungen nicht-menschlicher Handlungsfähigkeit analysiert.<sup>36</sup> Wie schon der Ausdruck „Aktanten“ deutlich macht, zeigen Latours Analysen, dass es keinesfalls um eine Anthropomorphisierung von digitalen Prozessen geht, sondern gerade umgekehrt um eine De-Antropomorphisierung von Softwareagenten, die – wie Informationswissenschaftler sagen – als „mindless machine(s)“<sup>37</sup> durch Handlungszurechnung zu nicht-menschlichen Mitgliedern der Gesellschaft geworden sind.

## 2. Kommunikation mit Aktanten

Doch können wir wirklich davon ausgehen, dass Algorithmen kommunizieren, wie wir das bei Menschen, aber auch bei Organisationen voraussetzen? Steckt in unserer Begegnung mit Algorithmen nur eine Wahrnehmung von Rechenergebnissen oder eine Kommunikation mit gesellschaftlichen Akteuren? Auch hier wieder in aller Kürze: Die Antwort hängt davon ab, ob es in der Kommunikation gelingt, die „Beiträge“ der Softwareagenten als kommunikative Ereignisse im strengen Sinne, also als Einheit von Information, Mitteilung und Verstehen zu aktivieren.<sup>38</sup> Sofern es im kommunikativen Prozess Mensch-Computer gelingt, Ereignisse zu identifizieren, die als „Mitteilungen“ der Algorithmen, welche eine bestimmte „Information“ enthalten, zu „verstehen“ sind, wird sich in der Tat in der Begegnung von Menschen mit Algorithmen ein genuines Sozialsystem herausbilden. Die „Antworten“ in der Form von Kommunikationen, die

the case of software agents, a computer is transformed into a *homo ex machina*. In both cases the stakes are the same: the social attribution of the ability to act to communication processes.

In his famous actor-network theory, in which the neologism “actants” is applied to digital processes capable of action, Bruno Latour formulated the fundamental difference between software agents’ and humans’ capacity for action and at the same time analyzed the more precise conditions of non-human ability to act.<sup>36</sup> As the term “actants” already makes clear, Latour’s analyses show that we are not dealing with anthropomorphizing digital processes, but just the other way round with de-anthropomorphizing software agents. They remain “mindless machines(s)”<sup>37</sup>, but via social attribution of action they become non-human members of society.

## 2. Communication with actants

But can we really assume that algorithms communicate, as we expect from people, but also from organizations? In our encounter with algorithms, is there only a perception of calculation results or a genuine communication with societal actors? Here again in a nutshell: The answer depends on whether communication succeeds in activating the “contributions” of software agents as communicative events in the strict sense, i.e. as a synthesis of utterance, information, and understanding.<sup>38</sup> If the communicative process succeeds in identifying events that are to be “understood” as “utterances” of the algorithms that contain a certain “information”, a genuine social system will indeed emerge in the encounter of human beings with algorithms. The “answers” in the form of communications we receive from software agents to our queries fulfill everything that the synthesis of utterance, information, and understanding, which is a prerequisite for

36 Bruno Latour, *Das Parlament der Dinge: Für eine politische Ökologie* (2001), 93 ff. Latour definiert ihre Handlungsfähigkeit sehr weit als „Widerständigkeit“. Demgegenüber wird im Text sehr viel enger auf ihre Teilnahme an gesellschaftlicher Kommunikation abgestellt.

37 Mireille Hildebrandt, *Smart Technologies and the End(s) of Law* (2015), 22; Floridi/Sanders (Fn. 6), 186.

38 Dazu grundlegend Luhmann (Fn. 32, 1984), 191 ff. In unserem Zusammenhang besonders wichtig ist die Möglichkeit der Kommunikation im strengen Sinne mit nicht-menschlichen Entitäten: Sprenger (Fn. 32), 116 ff.; Luhmann (Fn. 32, 1984), 116 ff. 119 ff.; speziell zur Kommunikation mit Algorithmen Elena Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, *Zeitschrift für Soziologie* 46 (2017), 249–265, 254 ff.; Teubner (Fn. 29).

36 Bruno Latour, *Politics of Nature: How to Bring the Sciences into Democracy* (2004), 62 ff. Latour broadly defines their ability to act as resistance. In contrast, the text focuses much more closely on their participation in social communication.

37 Mireille Hildebrandt, *Smart Technologies and the End(s) of Law* (2015), 22; Floridi/Sanders (fn. 6), 186.

38 See Luhmann (fn. 32, 1995), 140 ff. In our context, the possibility of communication in a strict sense with non-human entities is particularly important, Sprenger (fn. 32), 116 ff.; Luhmann (fn. 32, 2012/2013), Ch. 1, VI; especially for communication with algorithms, Elena Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, *Zeitschrift für Soziologie* 46 (2017), 249–265, 254 ff.; Teubner (fn. 29).

wir von Softwareagenten auf unsere Anfragen erhalten, erfüllen alles, was die Synthese von Information, Mitteilung und Verstehen, die Voraussetzung für Kommunikation im strengen Sinne ist, erfordert.<sup>39</sup> Und dies gilt auch, wenngleich schwieriger zu begründen, in der Gegenrichtung, in der Kommunikation von Mensch zu Computer.

Zwar ist dies eine „asymmetrische“ Interaktion von Menschen und Algorithmen, dennoch bildet sich zwischen ihnen ein genuines selbstproduzierendes Kommunikationssystem heraus. Sie ist asymmetrisch in einem dreifachen Sinne.

(1) Den Algorithmen können keineswegs die psychischen Kompetenzen von Menschen zugebilligt werden.<sup>40</sup> Ihr Innenleben besteht aus mathematischen Operationen auf der Grundlage elektronischer Signale. Aber dieser Unterschied der Innerlichkeit von Mensch und Computer ist für unsere Frage gar nicht relevant, sofern nur außerhalb der unterschiedlichen Innenleben von Mensch und Computer ein Kommunikationsgeschehen zustande kommt und zugleich das elektronische Innenleben der Algorithmen und das Bewusstsein der Menschen die Kommunikation zwischen ihnen ausreichend irritieren, sodass die Synthese aus Mitteilung, Information, Verstehen gelingen kann.

(2) Sie ist in einem anderen Sinne asymmetrisch. In der Kommunikation unter Menschen ist die doppelte Kontingenz symmetrisch beidseitig, weil beide Partner die Wahl ihres Verhaltens von der Verhaltenswahl des anderen abhängig machen.<sup>41</sup> Dagegen wird zwischen Mensch und Maschine die für Kommunikation notwendige doppelte Kontingenz nur einseitig erlebt, nämlich nur vom Menschen und nicht von der Maschine (jedenfalls im heutigen Entwicklungsstand).<sup>42</sup> Einseitig erlebte doppelte Kontingenz, wie wir sie in der Mensch-Maschinen-Beziehung antreffen, schließt aber nicht aus, dass dennoch Kommunikation stattfindet. Historisch bekannte Konfigurationen, wie die Kommunikation mit Gott im Gebet ebenso wie animistische Praktiken, leisten durchaus die Synthese aus Mitteilung, Information und Verstehen.<sup>43</sup> Aber sie tun dies nur unter der Voraussetzung, dass im Ereignisstrom eine Personifikation des nicht-menschlichen Partners stattfindet, welche eine Zurechnung der Handlung auf den „anderen“ ermöglicht. „Personalität entsteht, wo immer das

communication in the strict sense, requires.<sup>39</sup> And this also applies, albeit more difficult to justify in the opposite direction, in communication from person to computer.

Although the interaction of humans and algorithms is “asymmetric”, a genuine self-producing communication system emerges between them. It is asymmetrical in a threefold sense.

(1) The operations of the algorithms cannot in any way be equated with the mental operations of humans.<sup>40</sup> Their inner workings consist of mathematical operations based on electronic signals. But this difference in the inwardness of humans and computers is not relevant to our question at all, provided that only outside of the different inner life of human and computer does a communication process take place, and at the same time the electronic inner life of the algorithms and the consciousness of humans and computers are sufficient to irritate the communication between them, so that the synthesis of utterance, information, and understanding can succeed.

(2) It is asymmetrical in another sense. In communication between people, double contingency is symmetrical on both sides, because both partners make the choice of their behavior dependent on the choice of behavior of the other.<sup>41</sup> On the other hand, in the communication between human and machine, the double contingency is experienced only one-sidedly, i.e. only by the human and not by the machine (at least in the current state of development).<sup>42</sup> But such a unilaterally experienced double contingency, as we find it in the human-machine relationship, does not rule out the possibility of communication. Historically known configurations, such as communication with God in prayer as well as animistic practices, do indeed provide a synthesis of utterance, information, and understanding.<sup>43</sup> But they do this only on the condition that a personification of the non-human partner takes place in the event stream, which enables the action to be attributed to the “other”. “Personality arises wherever the behavior of others is imagined as a genuine choice and can be influenced communica-

39 Zur Theoriefrage, ob Arbeit mit Computern als Kommunikation zu verstehen ist, auch wenn doppelte Kontingenz nur einseitig erfahren wird, *Luhmann* (Fn. 32, 1997), 303 f. *Esposito* (Fn. 38), 262 stellt die Frage, ob der Kommunikationsbegriff erweitert oder ein anderer Begriff für die Begegnung Computer-Mensch entwickelt werden muss, lässt die Antwort dann aber letztlich offen.

40 *Esposito* (Fn. 38), 250.

41 Zu doppelter Kontingenz *Talcott Parsons/Edward A. Shils*, *Toward a General Theory of Action: Theoretical Foundations for the Social Sciences* (1951); *Luhmann* (Fn. 32, 1984), 148 ff.

42 Dazu *Luhmann* (Fn. 32, 1997), 304.

43 *Sprenger* (Fn. 32), 119 ff.

39 On the question of whether working with computers is to be understood as communication, even if double contingency is experienced only one-sidedly, *Luhmann* (fn. 32, 2012/2013), Ch. 1, VI. *Esposito* (fn. 38), 262 asks whether the concept of communication has to be expanded or another term for the encounter between computer and humans has to be developed and leaves the question open.

40 *Esposito* (fn. 38), 250.

41 On the fundamental concept of double contingency, *Talcott Parsons/Edward A. Shils*, *Toward a General Theory of Action: Theoretical Foundations for the Social Sciences* (1951); *Luhmann* (fn. 32, 1984), 148 ff.

42 See *Luhmann* (fn. 32, 2012/2013), Ch. 2, VII.

43 *Sprenger* (fn. 32), 119 ff.

Verhalten anderer als gewählt vorgestellt wird und durch eigenes Verhalten kommunikativ zu beeinflussen [ist].“<sup>44</sup> Die Personifizierung nicht-menschlicher Entitäten als ein performatives Ereignis, das die Person als ein semantisches Konstrukt konstituiert, kompensiert also diese Asymmetrie in der Mensch-Maschinen-Beziehung.

(3) Sie ist schließlich asymmetrisch in Bezug auf das wechselseitige Verstehen von Mensch und Maschine. Wenn Verstehen verstanden wird als die Fähigkeit, die Selbstreferenz des anderen in der eigenen Selbstreferenz nachzukonstruieren, dann können Menschen die Botschaften des Algorithmus verstehen, während dem Algorithmus die Fähigkeit des Verstehens des menschlichen Innenlebens fehlen dürfte.

In unserem Zusammenhang kann man die Frage offenlassen, weil es in der Kommunikation auf dieses wechselseitige Tiefenverstehen gar nicht ankommt. Man muss deutlich zwischen Verstehensprozessen innerhalb der Kommunikation und Verstehensprozessen innerhalb der interagierenden Entitäten unterscheiden.<sup>45</sup> Für Verstehen innerhalb des Kommunikationsprozesses ist nicht entscheidend, ob der Algorithmus selbst versteht, sondern ob der „Antworttext“ des Algorithmus den „Fragentext“ des Menschen versteht. Sofern die Kommunikation des Algorithmus die Differenz von Mitteilung und Information in der Kommunikation des Menschen nachvollzieht und darauf mit einer eigenen Differenz von Mitteilung und Information reagiert, ist ein kommunikativer Verstehensakt vollzogen, ohne dass es – um es nochmals zu betonen – darauf ankäme, ob im Innenleben des Algorithmus die Äußerung des Menschen verstanden wird. Auch hier bieten die Kommunikation mit Gott und animistische Praktiken historische Belege für kommunikatives Verstehen, das auch dann zustande kommt, wenn der „andere“ nicht versteht.

Um das bisher Gesagte in einer Kurzformel zusammenzufassen: Softwareagenten sind – ebenso wie Unternehmen und andere formale Organisationen – nichts anderes als bloße Informationsströme, die dann zu „Personen“ (oder Teilpersonen) werden, wenn sie im Kommunikationsprozess eine soziale Identität aufbauen und wenn ihnen zusammen mit den notwendigen organisatorischen Vorkehrungen, z. B. Vertretungsregeln, eigene Handlungsfähigkeit effektiv zugerechnet wird.

tively by one’s own behavior.”<sup>44</sup> The personification of non-human entities is a performative event that constitutes the person as a semantic construct and compensates for the asymmetry in the human-machine relationship.

(3) It is ultimately asymmetrical in relation to the mutual understanding of human and machine. If understanding is defined as the ability to reconstruct the other’s self-reference in one’s own self-reference, then humans can understand the messages of the algorithm, while the algorithm may lack the ability to understand the human inner life.

However, in our context, it is possible to leave the question unanswered, because such a mutual deep understanding is not at all important for communication. A clear distinction must be made between understanding within the communication process and understanding within the inner life of the interacting entities.<sup>45</sup> For understanding within the communication process it is not important whether the algorithm understands the human’s intentions, but whether the “answer text” of the algorithm understands the “question text” of the human being. Provided that the algorithm’s communication comprehends the difference between utterance and information in the human’s communication and reacts to it with its own difference of utterance and information, a communicative act of understanding has been carried out without – to emphasize it once again – depending on whether the inner workings of the algorithm understands the human’s intentions. Here, too, communication with God and animistic practices provide historical evidence of a communicative understanding, which comes about even if the “other” does not understand the human’s inner life.

To summarize what has been said so far in a short formula: Software agents – just like companies and other formal organizations – are nothing more than mere streams of information that become “persons” (or sub-persons) when they build up a social identity in the communication process and when they are effectively attributed with the ability to act, together with the necessary organizational arrangements, e.g. rules of representation.

<sup>44</sup> Luhmann (Fn. 32, 1984), 643.

<sup>45</sup> Niklas Luhmann, *Systeme verstehen Systeme*, in: Luhmann/Schorr (Hg.), *Zwischen Intransparenz und Verstehen: Fragen an die Pädagogik* (1986), 72–117, 93 ff.

<sup>44</sup> Luhmann (fn. 32, 2012/2013), Ch. 4, IV.

<sup>45</sup> Niklas Luhmann, *Systeme verstehen Systeme*, in: Luhmann/Schorr (ed.), *Zwischen Intransparenz und Verstehen: Fragen an die Pädagogik* (1986), 72–117, 93 ff.

### 3. Gradualisierte digitale Autonomie

Ob nun ein Softwareagent, also ein konkreter digitaler Informationsstrom, darüber hinaus als autonomer Akteur zu qualifizieren ist, ist dann die juristisch entscheidende Frage. Dies hängt nicht nur von der ihm zugeschriebenen sozialen Handlungsfähigkeit ab, sondern zusätzlich davon, mit welchen besonderen Qualitäten er in unterschiedlichen Kontexten als eigenständige Person ausgestattet wird. Jeder Sozialkontext schafft sich seine jeweils eigenen Kriterien der Personalität, die Wirtschaft nicht anders als die Politik, die Wissenschaft, die Moralphilosophie – oder das Recht.<sup>46</sup> Jedes Subsystem schreibt auf eine andere Weise Individuen, Kollektiven oder Algorithmen als seinen „Personen“ Handlungen, Entscheidungen, Vermögen, Verantwortlichkeiten, Rechte und Pflichten zu oder stattet sie mit Kapital, Interessen, Intentionen, Handlungszielen oder Präferenzen aus. Je nach gesellschaftlichem Kontext stößt man auf deutliche Variationen der effektiv geltenden Akteurskonzepte, wie an den unterschiedlichen Definitionen des *homo oeconomicus*, *homo juridicus*, *politicus* etc. deutlich wird. Und auch bei Softwareagenten werden die sozialen Kompetenzen vom jeweiligen Kommunikationssystem determiniert, weil es die soziale Identität des Algorithmus konstituiert, wie auch dessen Fähigkeit zu handeln, zu kommunizieren, zu entscheiden festlegt. Nur ein Beispiel: Soziale Bewegungen sind in der Politik als eigenständige Kollektivakteure anerkannt, in der Wirtschaft und im Recht gelten sie als Nicht-Personen.

In der interdisziplinären Diskussion werden nun ganz unterschiedliche Kriterien angeboten, ab wann einem Softwareagenten Autonomie zuzusprechen ist.<sup>47</sup> In vielen Disziplinen wird die Frage, ob Softwareagenten autonom handeln können, positiv beantwortet. Der Schwellenwert aber, von dem ab Autonomie zugeschrieben werden kann, ist kontrovers. Digitale Autonomie scheint ein gradualisiertes Phänomen zu sein.<sup>48</sup> Und die Gradualisierung findet nicht nur auf einer einzigen Skala mit unterschiedlichen Autonomiegraden statt, sondern in einem mehrdimensionalen Raum, der mehr oder weniger Autonomie zulässt.<sup>49</sup>

So legt der Oxforder Informationsphilosoph Floridi in drei Dimensionen Schwellenwerte für die Zurechnung von Handlungsfähigkeit an nicht-menschliche Entitäten, sowohl an Organisationen

<sup>46</sup> Sprenger (Fn. 32), 119 ff.

<sup>47</sup> Einen Überblick bietet Zech (Fn. 3), 170 f., Fn. 16.

<sup>48</sup> Hildebrandt (Fn. 37), 21.

<sup>49</sup> So besonders die mehrdimensionalen Kriterienkataloge, Lohmann (Fn. 49), 155; Zech (Fn. 3), 176; Catrin Misselhorn, Collective Agency and Cooperation in Natural and Artificial Systems, in: Misselhorn (Hg.), Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation (2015), 3–24, 5 ff.; Thürmel (Fn. 35), 53 ff.; Floridi/Sanders (Fn. 6), 192 f.

### 3. Gradualized digital autonomy

Whether a software agent, i.e. a concrete flow of digital information, can be qualified as an autonomous actor is then the crucial question for the law. This not only depends on the social capacity for action attributed to it, but also on the special qualities with which it is endowed as an independent person in different contexts. Each social context creates its own individual criteria of personality, the economy no different from politics, science, moral philosophy – or law.<sup>46</sup> Each subsystem attributes actions, decisions, assets, responsibilities, rights and obligations in a different way to individuals, collectives or algorithms as its “persons”, or equips them with capital, interests, intentions, goals or preferences. Depending on the societal context, there are clear variations in the concepts of actors, as can be seen from the different definitions of *homo oeconomicus*, *homo juridicus*, *politicus*, etc. And also in the case of software agents, the social competences are determined by the respective communication system, because it constitutes the social identity of the algorithm and determines its ability to act, to communicate, to decide. Just one example: Social movements are recognized in politics as independent collective actors; in the economy and in the law they are regarded as non-persons.

In the interdisciplinary discussion, quite different criteria are now offered as the starting point, from which a software agent can be granted autonomy.<sup>47</sup> In many disciplines, the question of whether software agents can act autonomously is answered positively. However, the threshold value from which autonomy can be attributed is controversial. Digital autonomy seems to be a gradualized phenomenon.<sup>48</sup> And gradualization does not only take place on a single scale with different degrees of autonomy, but in a multidimensional space that allows more or less autonomy.<sup>49</sup>

For example, the information philosopher Floridi sets thresholds in three dimensions for the attribution of the ability to act for non-human entities, both in organizations and algorithms: (1) interac-

<sup>46</sup> Sprenger (fn. 32), 119 ff.

<sup>47</sup> For an overview, see Zech (fn. 3), 170 f., fn. 16.

<sup>48</sup> Hildebrandt (fn. 37), 21.

<sup>49</sup> See especially the multi-dimensional criteria catalogues, Lohmann (fn. 49), 155; Zech (fn. 3), 176; Catrin Misselhorn, Collective Agency and Cooperation in Natural and Artificial Systems, in: Misselhorn (ed.), Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation (2015), 3–24, 5 ff.; Thürmel (fn. 35), 53 ff.; Floridi/Sanders (fn. 6), 192 f.



als auch an Algorithmen, fest: (1) Interaktion (mit Mitarbeitern und anderen Organisationen), (2) Fähigkeit, aus sich selbst Zustands-änderungen herbeizuführen und (3) Anpassung der Strategien für Entscheidungen. Andere wiederum stellen auf ganz heterogene Eigenschaften ab: auf Denkfähigkeit, Kommunikationsfähigkeit, Verstehen, rationales Handeln, Nicht-Prognostizierbarkeit ihrer Konditionalprogramme,<sup>50</sup> autonome Raumveränderung ohne menschliche Aufsicht,<sup>51</sup> geringe Strukturiertheit des Einsatzbereichs,<sup>52</sup> eigene Zielverfolgung und Mittelwahl,<sup>53</sup> Optimierung mehrerer Ziele,<sup>54</sup> Kontrollfähigkeit, Programmierfähigkeit, Lernfähigkeit,<sup>55</sup> Selbsteränderung, Selbstbewusstsein,<sup>56</sup> Künstliche Intelligenz, moralische Selbstgesetzgebung oder gar Leidensfähigkeit oder schließlich auf ein digitales Gewissen ab.<sup>57</sup>

Die Unterschiede gehen nicht notwendig auf wissenschaftliche Kontroversen zurück, die zugunsten der einen richtigen Lösung zu entscheiden wären, sondern erklären sich häufig aus dem jeweiligen Erkenntnisinteresse der beteiligten Disziplinen sowie aus praktischen Handlungsorientierungen in verschiedenen gesellschaftlichen Bereichen. Während die an Erklärung und Voraussage interessierten Kausalwissenschaften nur dann von Autonomie sprechen, wenn sie eine *black box* modellieren, innerhalb derer sie keine Kausalzusammenhänge mehr analysieren können, wohl aber ihr Außenverhalten beobachten, bauen die sinnverstehenden Sozialwissenschaften konstitutiv auf der Autonomie der Akteure auf, aber auch hier wieder mit deutlichen Unterschieden. So stellt die Ökonomie nutzenorientierte Entscheidungen heraus, definiert also Autonomie als *rational choice*, während Moral und Ethik Autonomie eher in Formen eines digitalen Selbstbewusstseins suchen.

#### 4. Rechtliche Autonomiekriterien

Das Rechtssystem wiederum muss aus eigenem disziplinären Erkenntnisinteresse und aus eigener Handlungsorientierung die rechtlich relevante Grenze zwischen instrumentalisiertem und autonomem Handeln selbst festlegen. Dabei wird es sich aber an der interdisziplinären Diskussion in den Informationswissenschaften orientieren und am Ende ein Autonomiekriterium wählen müssen,

tion (with employees and other organizations), (2) the ability to effect changes of state from within oneself, and (3) adaptation of strategies for decisions. Others, in turn, focus on quite heterogeneous properties: on the ability to think, communicate, understand, act rationally, on non-predictability of their conditional programs<sup>50</sup>, autonomous spatial change without human supervision<sup>51</sup>, low degree of structuring of the area of application,<sup>52</sup> pursual of proper aims and choice of means,<sup>53</sup> optimization of multiple aims,<sup>54</sup> control ability, programming ability, ability to learn,<sup>55</sup> self-change, self-consciousness,<sup>56</sup> artificial intelligence, moral self-regulation or even on the capacity to suffer, or ultimately on digital conscience.<sup>57</sup>

The differences do not necessarily stem from controversies, which would have to be decided in favor of the one right solution, but often are cleared up through the respective interest in knowledge of the participating disciplines as well as from practical action orientations in various social areas. While the causal sciences interested in explanation and prediction speak of autonomy only if they model a black box in which they can no longer analyze causal relationships but observe their external behavior, the interactionist social sciences rely on constitutive autonomy of the actors – but here again with clear differences as well. Economics highlights utility-oriented decisions, defining autonomy as rational choice, while morality and ethics tend to seek autonomy in the form of a digital consciousness.

#### 4. Legal criteria of autonomy

The legal system, in turn, must define the legally relevant borderline between instrumentalized and autonomous action on the basis of its own disciplinary interest in knowledge and its own focus on action. In so doing, however, it must orient itself on the interdisciplinary discussion within information sciences and ultimately choose an autonomy criterion that is justifiable in line with the interdis-

50 Zech (Fn. 3), 171 f.; Bräutigam/Klindt (Fn. 12), 1137.

51 Matthias (Fn. 4), 35.

52 Lohmann (Fn. 49), 154.

53 Misselhorn (Fn. 49), 6 f.

54 Karanasiou/Pinotsi (Fn. 26), 1.

55 Matthias (Fn. 4), 17 ff.

56 Zimmerman (Fn. 17), 34 ff.

57 Eine informative Diskussion dieser unterschiedlichen Kriterien bei Misselhorn (Fn. 49), 4 ff.

50 Zech (fn. 3), 171 f.; Bräutigam/Klindt (fn. 12), 1137.

51 Matthias (fn. 4), 35.

52 Lohmann (fn. 49), 154.

53 Misselhorn (fn. 49), 6 f.

54 Karanasiou/Pinotsi (fn. 26), 1.

55 Matthias (fn. 4), 17 ff.

56 Zimmerman (fn. 17), 34 ff.

57 For an informative discussion of these different criteria, Misselhorn (fn. 49), 4 ff.

das vor dem interdisziplinären Diskussionsstand vertretbar ist.<sup>58</sup> Ähnlich wie das Recht auch sonst, etwa in Umweltfragen, vorgeht, wenn es angesichts von wissenschaftlichen Befunden einer Gradualisierung selbstverantwortlich Grenzwerte festlegt, so muss es hier aufgrund rechtseigener Kriterien festlegen, ab welchem Grad von wissenschaftlich festgestellter Eigenständigkeit digitaler Phänomene im Rechtssinne von deren Autonomie auszugehen ist.<sup>59</sup>

Denkfähigkeit oder künstliche Intelligenz wird in der Debatte immer wieder als das eigentliche Kriterium vorgeschlagen, das über Rechtssubjektivität entscheidet.<sup>60</sup> Hier nun ist eine verbreitete Fehlvorstellung zu korrigieren. Für ihre Handlungsfähigkeit geht es gar nicht um die Frage: Welche Art von ontologischen Eigenschaften – Intelligenz, Geist, Seele, reflexive Kapazitäten, Einfühlungsvermögen – muss ein Softwareagent besitzen, um als Akteur gelten zu können?<sup>61</sup> Auch hier wieder ist das Paradigma von formalen Organisationen als juristischen Personen hilfreich: Für die rechtliche Handlungsfähigkeit nicht-menschlicher Agenten sind innere psychische Zustände nicht ausschlaggebend.<sup>62</sup> „After all, what is interesting in the interaction with algorithms is not what happens in the machine’s artificial brain, but what the machine tells its users and the consequences of this.“<sup>63</sup> Nicht die inneren Eigenschaften der Agenten, sondern die gesellschaftliche Interaktionen, an der die laufenden Operationen des Algorithmus teilnehmen, d. h. eine geschlossene und autonome Abfolge rekursiver Kommunikationen, konstituieren den Algorithmus als Person, als semantisches Artefakt, als kommunikationsfähigen Akteur.<sup>64</sup> Die Operationen des Algorithmus, also digitale Datenströme, existieren – wie oben schon ausgeführt – als Akteure nicht per se, sondern sie werden erst von Sozialsystemen als adressierfähige Personen konstituiert. Gesellschaftliche Teilsysteme konstruieren sie als semantische Artefakte, indem sie ihnen volle oder eingeschränkte Subjektivität zuschreiben.

ciplinary state of the discussion.<sup>58</sup> Similarly to environmental matters, when the law defines threshold values for liability for damages in view of a scientifically determined scale of environmental degradation, it must determine, on the basis of its own legal criteria, at what degree digital phenomena can be assumed to be autonomous in the legal sense.<sup>59</sup>

Capacity for artificial intelligence is repeatedly proposed in the debate as the decisive criterion that determines autonomy and thus legal subjectivity.<sup>60</sup> Here a common misconception has to be corrected. Their ability to act depends not at all on the question: What kind of ontological characteristics – intelligence, mind, soul, reflexive capacities, empathy – does a software agent have to possess in order to be considered an actor?<sup>61</sup> Here again, the paradigm of formal organizations as legal entities is helpful: for the legal capacity of non-human agents, inner psychic states are not decisive.<sup>62</sup> “After all, what is interesting in the interaction with algorithms is not what happens in the machine’s artificial brain, but what the machine tells its users and the consequences of this.”<sup>63</sup> Not the inner properties of the agents, but the social interaction in which the current operations of the algorithm participate, i.e. a closed and autonomous sequence of recursive communications, constitute the algorithm as a person, as a semantic artifact, as a communicative actor.<sup>64</sup> As already mentioned above, the algorithms do not exist as actors per se, but are constituted by social systems as addressable persons. Social subsystems construct them as semantic artifacts by ascribing full or limited subjectivity to them.

58 Allgemein zur Rolle des Rechts in interdisziplinären Zusammenhängen *Gunther Teubner*, Rechtswissenschaft und -praxis im Kontext der Sozialtheorie, in: Grundmann/Thiessen (Hg.), *Recht und Sozialtheorie im Rechtsvergleich: Interdisziplinäres Denken in Rechtswissenschaft und -praxis* (2015), 145–168.

59 *Matthias* (Fn. 4), 43 ff.

60 Z. B. *Spindler* (Fn. 12), 816.

61 So entschieden gegen einen Trend in den Ingenieurwissenschaften, die gegenüber den „inneren“ Prozessen der Algorithmen die sozialen Interaktionen vernachlässigen, *Esposito* (Fn. 39), 250; ebenso *Latour* (Fn. 36), 93 ff.; *Teubner* (Fn. 29).

62 *Niklas Luhmann*, *Organisation und Entscheidung* (2000), Kap. 13 IV.

63 *Esposito* (Fn. 39), 250.

64 Für die Konstitution als Person aus philosophischer Sicht *Daniel Dennett*, *The Intentional Stance* (1987), 17; aus systemtheoretischer Sicht *Luhmann*, *Nachweise in* (Fn. 32) (nur Kollektivakteur) und *Sprengrer* (Fn. 32), 114. Speziell für Softwareagenten, *Gruber* (Fn. 28, 2015), 250 ff.; *Matthias* (Fn. 4), 83 ff.; *Teubner* (Fn. 29); *Solum* (Fn. 17).

58 In general on the role of law in interdisciplinary contexts, *Gunther Teubner*, *Law and Social Theory: Three Problems*, in: *Ancilla Juris* (anci.ch) 2014, 182–221.

59 *Matthias* (fn. 4), 43 ff.

60 E.g. *Spindler* (fn. 12), 816.

61 In this sense, against a trend in engineering sciences that neglects social interactions as compared to the “inner” processes of algorithms, *Esposito* (fn. 39), 250; also *Latour* (fn. 36), 62 ff.; *Teubner* (fn. 29).

62 *Niklas Luhmann*, *Organisation und Entscheidung* (2000), Kap. 13 IV.

63 *Esposito* (fn. 39), 250.

64 For the constitution as a person from a cognitive science point of view, *Daniel Dennett*, *The Intentional Stance* (1987), 17; from a systems theory point of view for collective actors, *Luhmann* (fn. 32) and for non-human actors, *Sprengrer* (fn. 32), 114. Especially for software agents, *Gruber* (fn. 28, 2015), 250 ff.; *Matthias* (fn. 4), 83 ff.; *Teubner* (fn. 29); *Solum* (fn. 17).

Zwar basiert dies auf der gesellschaftlichen Unterstellung, dass die kommunizierende Einheit kommunikative Fähigkeiten hat, aber der fiktionale Charakter ist kein Schaden, solange es nur gelingt, über ihre Beiträge den Kommunikationsfluss mit Erfolg aufrechtzuerhalten. Für den kommunikativen Umgang mit Personen ist ein Name erforderlich, nicht aber eine Entschlüsselung innerer Prozesse ‚in‘ der Person. Das gilt für Organisationen ebenso wie für Algorithmen. Also: Nicht auf die innere Denkfähigkeit der Algorithmen kommt es an, nicht auf „wirkliche“ künstliche Intelligenz, was immer das bedeutet, sondern auf ihre Beteiligung an gesellschaftlicher Kommunikation. „*Artificial communication*“ und nicht „*artificial intelligence*“ ist entscheidend.<sup>65</sup> Diese Entpsychologisierung, wie sie die Kommunikationstheorie vorschlägt, trifft sich recht genau mit den bekannten Objektivierungstendenzen im Recht der Willenserklärung und im Begriff der Fahrlässigkeit.<sup>66</sup> Darauf ist zurückzukommen.

Intentionales Handeln dagegen dürfte eine notwendige Voraussetzung für Autonomie sein, sofern darunter kein innerer psychischer Zustand verstanden wird, sondern die Zuschreibung zielgerichteten Handelns durch einen Beobachter gemeint ist – die berühmte „*intentional stance*“ des Kognitionswissenschaftlers Daniel Dennett: Ob die Agenten in Wahrheit Willensfreiheit besitzen, ist keine wissenschaftlich sinnvolle Frage. Stattdessen: Wenn wegen erhöhter Komplexität eine physikalische Beschreibung nicht möglich ist, kann die Wissenschaft mit Hilfe eines intentionalen Vokabulars das untersuchte System als einen Akteur, der mit Annahmen über die Welt, mit Zielen und Handlungsoptionen agiert, analysieren und dadurch neue Erkenntnisse gewinnen.<sup>67</sup> Die Systemtheorie weitet dies von der Wissenschaft auf andere Beobachter aus.<sup>68</sup> Nicht nur die Wissenschaft kann Softwareagenten intentional – zum Beispiel in einer Vertragsbeziehung – interpretieren, auch der Partner in einer Interaktion (Ego) kann Alter nicht mehr kausalistisch, sondern intentional interpretieren und dadurch neue Orientierungen für seine Handlungen finden. Ebenso kann ein ganzes Sozialsystem – in unserem Fall das Recht – dieser Beobachter sein, der Softwareagenten Intentionen zurechnet und daraus Konsequenzen für rechtliche Verbindlichkeit von dessen Erklärungen und Verantwortung für dessen Handlungen zieht.

Although this is based on the assumption that the communicating unit has action abilities, the fictional character is no flaw as long as it only succeeds in continuing the flow of communication through its contributions. In order to communicate with persons, a name is required, but not the decoding of inner processes “inside” the person. This applies to organizations as well as to algorithms. So: it is not the inner capacity for thought of the algorithms that is important, not “true” artificial intelligence, whatever that means, but their participation in social communication. “Artificial communication” and not “artificial intelligence” is crucial.<sup>65</sup> This depsychologization, as suggested by communication theory, is quite closely related to the known tendencies towards objectivization in the law of the declaration of intent and in the concept of negligence.<sup>66</sup> That is what we have to come back to.

Intentional action, on the other hand, is likely to be a necessary prerequisite for autonomy, provided that this does not mean an inner psychological state, but the attribution of purposeful action by an observer – the famous “intentional stance” proposed by cognitive scientist Daniel Dennett: Whether or not the agents actually possess freedom of will is not a scientifically meaningful question. Instead: if a physical description is not possible due to increased complexity, science can use an intentional vocabulary to analyze the investigated system as an actor who operates with assumptions about the world, with goals and options for action, and thus gain new insights.<sup>67</sup> Systems theory extends this from science to other observers.<sup>68</sup> It is not only science that can observe software agents as intentional actors, e.g. as party in a contractual relationship, but also the partner in an interaction (ego), who no longer interprets alter’s behavior in a causal but in an intentional manner and thus finds new orientations for his own actions. Similarly, an entire social system – in our case the law – can be this observer, who assigns intentions to software agents and draws consequences for the legally binding nature of their declarations and for the responsibility for their actions.

65 *Esposito* (Fn. 39), 250.

66 Dazu Münchener Kommentar BGB/Grundmann, 7. Aufl. 2016, § 276 Rn. 53 ff.

67 *Dennett* (Fn. 64), 17; *Matthias* (Fn. 4), 41 ff.

68 *Luhmann* (Fn. 32, 1984), 155 ff.

65 *Esposito* (fn. 39), 250.

66 See Münchener Kommentar BGB/Grundmann, 7th ed. 2016, BGB § 276 Rn. 53 ff.

67 *Dennett* (fn. 64), 17; *Matthias* (fn. 4), 41 ff.

68 *Luhmann* (fn. 32, 1995), 109 ff.

Jedoch für eine rechtlich relevante Autonomie dürfte die bloße Intentionalität, also Zielgerichtetheit und Mittelwahl eines Agenten, noch nicht ausreichen. Das gleiche gilt für dessen Teilnahme an Kommunikation. Denn auch bloß automatisierte Softwareagenten nehmen schon an Kommunikation teil. Ebenso wie Intentionalität ist die Teilnahme an Kommunikation nur eine notwendige, aber nicht hinreichende Bedingung.

Andere Kriterien wiederum dürften über die juristischen Minimalvoraussetzungen für Autonomie weit hinausgehen. Rationales Handeln etwa, wie es Dennett über die „*intentional stance*“ hinaus als Autonomievoraussetzung nicht-menschlicher Agenten verlangt, mag für *rational choice* von Wirtschaftsakteuren plausibel sein, nicht aber für Rechtsakteure, deren irrationales Handeln bei Rechtsverletzungen gerade im Vordergrund steht. Ebenso dürften auch andere anspruchsvolle Aktivitäten wie Selbständerung oder selbsttätiges Lernen schon die Minimalvoraussetzungen überschreiten. Gewiss steigern sie noch einmal erheblich die Handlungsautonomie der Agenten und ermöglichen es, Sanktionen und Anreize direkt auf die Agenten auszurichten.<sup>69</sup> Aber es dürfte unter Gesichtspunkten des Schadensausgleichs nicht angemessen sein, erst dann mit rechtlichen Verantwortungsmechanismen einzusetzen, wenn die Programmkorrekturen nur von ihnen selbst, nicht aber von Programmierern vorgenommen werden.

Schon gar nicht dürfen, wie schon gesagt, für eine rechtlich relevante Digitalautonomie eine vollausgebildete Künstliche Intelligenz, Empathie, Gefühle, Leidenschaft, Selbstbewusstsein oder gar ein digitales Gewissen verlangt werden. Erst recht kann ein haftungsrechtlich relevanter Autonomiebegriff für digitale Phänomene nicht Anleihen bei der philosophischen Tradition machen, wonach Autonomie als die Möglichkeit des Menschen verstanden wird, sich selbst als freiheits- und vernunftfähiges Wesen zu bestimmen und entsprechend aus Freiheit moralisch zu handeln.<sup>70</sup> Es ist unhaltbar zu behaupten, dass, erst wenn digitale Autonomie zu Selbstbewusstsein erstarkt ist, dann auch Rechtspersönlichkeit indiziert sei. Gewiss sind dies Fragen, die sich der Informationsphilosophie in ihrer Suche nach einer möglichen digitalen Moralität stellen. Aber wenn solche Eigenschaften im Haftungsrecht als Autonomiekriterium verlangt werden, dann leistet dies nur opportunistischem

However, for a legally relevant autonomy the mere intentionality, i.e. goal orientation and choice of means of the agent, is not sufficient. The same applies to its participation in communication. After all, even automated software agents are already taking part in communication. Just like intentionality, participation in communication is only a necessary but not sufficient condition.

Other criteria, in turn, are likely to go far beyond the minimum requirements for legal autonomy. Rational action, for example, as Dennett demands beyond “intentional stance” as a prerequisite for the autonomy of non-human agents, may be plausible for a rational choice by economic actors, but not for legal actors whose irrational action in the event of infringements is of particular importance. Similarly, other demanding activities such as self-alteration or self-learning are likely to exceed the minimum requirements. Certainly, they increase the degree of autonomy of the agents and make it possible to focus sanctions and incentives directly on the agents.<sup>69</sup> However, from the point of view of compensation for damages, it would not be appropriate to impose liability exclusively in those cases if the program corrections are only made by the algorithms themselves and not by programmers.

As I have already said, for a legally relevant digital autonomy it is not necessary to demand artificial intelligence, empathy, feelings, suffering, self-confidence or even a digital conscience. Even more so, a relevant concept of autonomy under liability law for digital phenomena cannot borrow from the philosophical tradition, according to which autonomy is understood as the self-determination of a person who is capable of freedom and reason and to act morally out of freedom.<sup>70</sup> It is unsustainable to claim that only when a digital agent develops self-consciousness, will legal personality be indicated. Certainly these are questions that are posed to the information philosophy in its search for a possible digital morality. But if such characteristics are demanded as a criterion for autonomy in liability law, then this would only encourage opportunistic behavior on the part of lawyers, who confidently stick to traditional doctrine today while simultaneously keeping a back door open in the event that

69 Floridi/Sanders (Fn. 6), 192 ff.

70 Immanuel Kant, Grundlegung zur Metaphysik der Sitten (2012), 60.

69 Floridi/Sanders (fn. 6), 192 ff.

70 Immanuel Kant, Grundlegung zur Metaphysik der Sitten (2012), 60.

Verhalten der Juristen Vorschub, die es für die Gegenwart getrost bei der traditionellen Dogmatik belassen, sich aber zugleich ein Hintertürchen offenhalten, für den Fall, dass sich die Juristen mit ihrer exklusiven Menschzurechnung einmal vollständig gesellschaftlich isolieren sollten.

##### 5. Autonome Entscheidungen der Softwareagenten

Entscheidung unter Ungewissheit – dies dürfte das rechtsrelevante Autonomiekriterium sein. Wenn diese Aufgabe an Softwareagenten delegiert wird und sie sich entsprechend verhalten, dann ist das Recht genötigt, ihnen Handlungsfähigkeit zuzusprechen. Softwareagenten handeln im Rechtsinne dann autonom, wenn ihr Verhalten nicht mehr ausschließlich einem Reiz-Reaktion-Schema folgt, sondern wenn sie eigene Ziele besitzen, abändern, verfolgen und nicht-prognostizierbare Entscheidungen treffen.<sup>71</sup>

Konkret bedeutet dies: Wenn (1) ein Softwareagent so programmiert ist, dass er zwischen Alternativen zu entscheiden hat, wenn (2) er diese Entscheidung als Optimierung verschiedener Kriterien treffen muss und wenn (3) ein Programmierer das Verhalten des Softwareagenten weder nachträglich erklären noch für die Zukunft voraussagen kann, sondern nur noch *ex post* korrigieren kann, dann sollte das Recht von Autonomie, d. h. von Entscheidungsfähigkeit und Verantwortungsfähigkeit des Softwareagenten ausgehen. Praktisch setzt dies die Verpflichtung für den Hersteller voraus, eine *black box* zu installieren, deren Protokollierungsfunktion den Entscheidungsablauf nachvollziehen lässt.<sup>72</sup> Jedoch – wie stets im Zwischenbereich zwischen technisch-wissenschaftlicher Expertise und Recht – bedeutet dies keine automatische Bindung an die Expertise, sondern bedarf einer eigenverantwortlichen Rechtsentscheidung. Vergleichbar der Beziehung zwischen Sachverständigen und Richter bei Fragen der Zurechnungsfähigkeit geht es darum, den Umschlagspunkt von Kausalzurechnung zu Entscheidungszurechnung im einzelnen zu bestimmen. Das Recht zieht bekanntlich in diesem Umschlag noch zusätzliche rechtspolitische und rechtsdogmatische Gesichtspunkte heran.

Warum ist Entscheidung unter Ungewissheit das rechtlich relevante Kriterium? Ungewissheit resultiert aus der Indeterminiertheit der Programmierung und einer geringen Strukturiertheit der Umwelt, mit welcher der Algorithmus konfrontiert wird.<sup>73</sup> Der Grund für die rechtliche Relevanz ist der fundamentale Zusammenhang von Entschei-

they, with their exclusive attribution of digital actions to humans, should one day completely isolate themselves socially.

##### 5. Autonomous decisions of software agents

Decision under uncertainty – this is likely to be the legally relevant criterion of autonomy. If such a decision is delegated to software agents and they behave accordingly, then the law is required to assign them legal action capacity. Software agents act autonomously in the legal sense, when their behavior no longer follows an exclusively stimulus-reaction scheme, but when they pursue their own goals and make decisions that cannot be predicted.<sup>71</sup>

In concrete terms, this means this: If (1) a software agent is programmed in such a way that it has to decide between alternatives, if (2) it has to make this decision as an optimization of various criteria, and if (3) a programmer cannot explain the behavior of the software agent retrospectively or predict it for the future, but can only correct it *ex post*, then the law should assume autonomy, i.e. the software agent's decision-making ability and responsibility. In practice, this implies an obligation for the manufacturer to install a black box, the logging function of which makes it possible to trace the decision process.<sup>72</sup> However, as always in the intermediate area between technical-scientific expertise and law, this does not automatically bind the law to the expertise, but requires a legal decision on its own authority. Comparable to the relationship in criminal law between experts and judges in matters of mental capacity, the aim is to determine in detail the point of transition from causal attribution to decision attribution. As is well known, law in this context takes account of additional aspects of legal doctrine and policy.

Why is decision under uncertainty the legally relevant criterion? Uncertainty results from the indeterminacy of programming and a low degree of structuring of the environment with which the algorithm is confronted.<sup>73</sup> The reason for its legal relevance is the fundamental connection between decision and responsibility.<sup>74</sup> There is an inextrica-

71 In dieser Richtung auch Europäisches Parlament (Fn. 3), 6; *Kirn/Müller-Hengstenberg* (Fn. 9), 4.

72 *Lohmann* (Fn. 49), 158; *Günther* (Fn. 17), 99.

73 *Lohmann* (Fn. 49), 154.

71 In this direction also the European Parliament (fn. 3), 6; *Kirn/Müller-Hengstenberg* (fn. 9), 4.

72 *Lohmann* (fn. 49), 158; *Günther* (fn. 17), 99.

73 *Lohmann* (fn. 49), 154.

dung und Verantwortung.<sup>74</sup> Zwischen der Öffnung von Entscheidungsalternativen in einer ungewissen Umwelt, deren Delegation an nicht-menschliche Prozessoren und der dabei entstehenden Verantwortungsproblematik besteht ein unauflöslicher Zusammenhang. Verantwortung im strengen Sinne ist das Einstehenmüssen für Entscheidung unter Ungewissheit, deren Resultat nicht prognostizierbar ist.<sup>75</sup> Es geht nicht bloß um das Einstehenmüssen für Fehler!<sup>76</sup> Wenn Softwareagenten bei voll determinierter Kalkulation Fehler machen, dann bedarf es einer bloßen Fehlerkorrektur. Anders aber bei nicht-determinierten Entscheidungen unter Ungewissheit. Wenn ein solches unkalkulierbares Risiko eingegangen wird, dann kann eine Fehlentscheidung vorher gar nicht vermieden werden, sondern, wenn sie trotz aller Vorkehrungen eintritt, nur noch bereut werden.<sup>77</sup> Dies nachträgliche Bereuen von Ungewissheitsentscheidungen aber ist ein klarer Fall von rechtlich gebotener Verantwortung, unter anderem auch von zivilrechtlicher Haftung.

Sollte man dann überhaupt das Risiko eingehen, dass Algorithmen nicht-prognostizierbare Entscheidungen fällen dürfen? Zech vertritt den strengen Standpunkt, dass schon nach geltendem Recht der Einsatz autonomer Algorithmen prinzipiell rechtswidrig ist. Nur der Gesetzgeber könne eine Ausnahme anordnen und dies auch nur dann, wenn er zugleich eine effektive Risikovorsorge trifft.<sup>78</sup> Ein solches Verbot mag als eine Extremlösung erscheinen, trifft aber das Problem genau: Das Risiko einer echten Entscheidungsdelegation an nicht-menschliche Akteure ist weder abschätzbar noch beherrschbar.<sup>79</sup>

Ungewissheitsentscheidungen mit Eigenrisiko sind gesellschaftlich sehr viel relevanter als reine Rechenaufgaben mit bloßem Fehlerrisiko. „Der Einsatz voll-autonomer beweglicher Maschinen im öffentlichen Raum dürfte am oberen Ende der Risikokala stehen.“<sup>80</sup> Auch der häufig angeführte Vorteil, die massive Senkung von Transaktionskosten, kann nicht aufwiegen, dies hohe Risiko einzugehen. Die eigentliche Rechtfertigung liegt im „Entde-

74 In soziologischer Sicht *Niklas Luhmann*, *Soziologie des Risikos* (1991), 197 ff.; aus der Sicht des Haftungsrechts *Nils Jansen*, *Die Struktur des Haftungsrechts: Geschichte, Theorie und Dogmatik außervertraglicher Ansprüche auf Schadensersatz* (2003), 136 ff.

75 Zum Zusammenhang von Entscheidungen unter Ungewissheit und Verantwortung *Niklas Luhmann*, *Funktionen und Folgen formaler Organisation* (1964), 172 ff.; *Niklas Luhmann*, *Ökologische Kommunikation: Kann die moderne Gesellschaft sich auf ökologische Gefährdungen einstellen?* (1986), 26.

76 „Only those questions that are in principle undecidable we can decide.“ Zu diesem Unterschied von Entscheidung und Errechnung *Heinz von Foerster*, *Ethics and Second-Order Cybernetics, Cybernetics and Human Knowing 1* (1992), 9–19.

77 Zum „postdecisional regret“ *Luhmann* (Fn. 62), 146, 170; *Luhmann* (Fn. 75), 32 f., 207 ff.

78 *Zech* (Fn. 3), 191 ff.

79 *Matthias* (Fn. 4), 33 ff.

80 *Zech* (Fn. 3), 176 setzt die Autonomieschwelle relativ hoch an, erst bei Selbständerung und Lernfähigkeit und mangelnder Vorausschbarkeit.

ble link between the opening up of decision alternatives in an uncertain environment, their delegation to non-human processors and the resulting problem of responsibility. Responsibility in the strict sense of the word is the obligation to take responsibility for decisions under uncertainty, the outcome of which cannot be predicted.<sup>75</sup> It's not just a question of answering for mistakes!<sup>76</sup> If software agents make mistakes with fully determined calculations, then a mere error correction is required. However, it is different in the case of undetermined decisions under uncertainty. If such an incalculable risk is taken, then a wrong decision cannot be avoided beforehand, but only regretted if it occurs despite all precautions.<sup>77</sup> This subsequent repentance of decisions under uncertainty, however, is a clear case of legally required responsibility, including private liability.

Should we then run the risk that algorithms are allowed to make decisions that cannot be predicted? Indeed, Herbert Zech takes the strict view that under current law the use of autonomous algorithms is in principle illegal. Only the legislature can order an exception and only if it simultaneously makes effective risk provisioning.<sup>78</sup> Such a ban may appear to be an extreme solution, but it points precisely to the problem: The risk of a genuine delegation of decisions to non-human actors is neither predictable nor controllable.<sup>79</sup>

Uncertainty decisions with inherent risk are much more relevant to society than purely mathematical tasks with a mere risk of error. “The use of fully-autonomous mobile machines in the public domain is likely to be at the top end of the risk scale.”<sup>80</sup> The frequently cited advantage, the massive reduction of transaction costs, cannot compensate for this high risk either. The real justification lies in the “discovery process” of the use of autonomous algo-

74 From a sociological point of view, *Niklas Luhmann*, *Risk: A Sociological Theory* (1993), ch. 10; from the point of view of liability law *Nils Jansen*, *Die Struktur des Haftungsrechts: Geschichte, Theorie und Dogmatik außervertraglicher Ansprüche auf Schadensersatz* (2003), 136 ff.

75 On the connection between uncertainty decisions and responsibility, *Niklas Luhmann*, *Funktionen und Folgen formaler Organisation* (1964), 172 ff.; *Niklas Luhmann*, *Ecological Communication* (1989), II.

76 “Only those questions that are in principle undecidable we can decide.“ On this difference between decision and calculation, *Heinz von Foerster*, *Ethics and Second-Order Cybernetics, Cybernetics and Human Knowing 1* (1992), 9–19.

77 On “postdecisional regret” *Luhmann* (fn. 62), 146, 170; *Luhmann* (fn. 74), Ch. 1, III; Ch. 10, II.

78 *Zech* (fn. 3), 191 ff.

79 *Matthias* (fn. 4), 33 ff.

80 *Zech* (fn. 3), 176 puts the threshold of autonomy relatively high, only when self-alteration, the ability to learn and lack of predictability are concerned.

ckungsverfahren“ des Einsatzes „autonomer Algorithmen“, in ihrem enormen Innovationspotential. Wenn Computer Entscheidungen unter Ungewissheit treffen, dann ermöglicht dies, etwas gänzlich Neues zu entdecken, etwas, das sich noch keine menschliche Intelligenz ausgedacht hat, und sogar manchmal, das keine menschliche Intelligenz nachvollziehen könnte, herauszufinden, und nicht einfach Errechenbares zu kalkulieren.<sup>81</sup>

Hier steckt das eigentliche Motiv, warum die Gesellschaft es zulässt, dass Menschen Entscheidungen unter Ungewissheit an Algorithmen delegieren. Damit nehmen sie das Risiko katastrophaler Fehlentscheidungen bewusst in Kauf, das Risiko, dass das Entdeckungsverfahren gesellschaftlich höchst unerwünschte Folgen zeitigt. Hier also liegt der tiefere Grund für gesteigerte Anforderung an Verantwortung für Entscheidungen im Unterschied der Verantwortung für bloße Rechenfehler. Ja, Verantwortung gewinnt ihre eigentliche Bedeutung erst hier: als Einstehenmüssen für den „Sprung ins Dunkle“.<sup>82</sup> Diesen Sprung ins Dunkle als Entscheidung unter Ungewissheit nicht nur Menschen anzuvertrauen, sondern Algorithmen zu überlassen, darin besteht das fundamental Neue. Wenn das Recht das Entdeckungsverfahren autonomer Algorithmen zulässt, wenn es erlaubt, dass Softwareagenten autonome Entscheidungen treffen dürfen, ist es unabdingbar, dass gerade dafür das Recht wirksame Verantwortungsformen für den Enttäuschungsfall bereitstellt. Und dies gilt gerade auch für den Fall, dass den beteiligten Menschen gar keine Pflichtverletzungen unterlaufen sind.

Digitale Ungewissheitsentscheidungen eröffnen ein gänzlich neues gesellschaftliches Labor für Experimente. Nur noch experimentell kann das Neue ausprobiert, nicht mehr vorherberechnet, sondern nur noch nachträglich auf seine Folgen überprüft werden.<sup>83</sup> Evolutionstheoretisch ausgedrückt: Digitale Ungewissheitsentscheidungen erzeugen vielfältige Variationen, auf die Menschen gar nicht gekommen wären. Verantwortungszuschreibung aber ermöglicht erst ihre gesellschaftlich vertretbare Selektion, die schadensstiftende Variationen wieder ausscheidet. Und in der auf Dauer gestellten Normativierung des richtigen Entscheidens liegt die Retention. Darf man – so muss man die Autoren fragen, welche die eingangs

rithms, in their enormous potential for innovation. When computers make decisions under uncertainty, it makes it possible to discover something completely new, something that has not yet been invented by human intelligence, and sometimes even something that no human intelligence can comprehend, and not merely a calculation of the calculable.<sup>81</sup>

This is the real reason why society allows people to delegate decisions to algorithms under uncertainty. In doing so, they consciously accept the risk of catastrophically wrong decisions, the risk that the discovery process will have highly undesirable consequences for society. This underlies the increased demand on responsibility for decisions, in contrast to responsibility for simple arithmetic errors. It is only here that responsibility takes on its true meaning: as a consequence of the “leap into the dark”.<sup>82</sup> To not only entrust this leap into the dark to people, but to leave them to algorithms is the fundamentally new thing. If the law allows for the discovery processes of autonomous algorithms, if it allows software agents to make autonomous decisions, it is essential that the law provides effective forms of responsibility in case of disappointment. And this applies especially in the event that the human actors involved have not committed any breaches of their obligations.

Digital uncertainty decisions open up a completely new social laboratory for experimentation. It is only by experiment that the action can be tried out, no longer calculated in advance, but only subsequently evaluated by its consequences.<sup>83</sup> In terms of evolutionary theory, digital decisions under uncertainty create a multitude of variations that humans would never have thought of. However, attribution of responsibility makes it possible to make socially justifiable selections, which in turn eliminate harmful variations. And in the permanent juridification of the decision lies retention. May we burden the injured party with these new risks as ordinary contingencies of life – we must ask the authors who are willing to accept the above-mentioned gaps in

81 *Esposito* (Fn. 39), 253.

82 Trotz aller Rationalisierung von Entscheidungen bleibt es beim „Mysterium“ der Entscheidung, *Niklas Luhmann*, Die Paradoxie des Entscheidens, *Verwaltungsarchiv* 84 (1993), 287–310, 288.

83 *Matthias* (Fn. 4), 33 ff.

81 *Esposito* (fn. 39), 253.

82 Despite all the rationalization, the “mystery” of the decision remains, *Niklas Luhmann*, Die Paradoxie des Entscheidens, *Verwaltungsarchiv* 84 (1993), 287–310, 288.

83 *Matthias* (fn. 4), 33 ff.

geschilderten Verantwortungslücken hinzunehmen bereit sind – diese neuartigen Risiken den Geschädigten als „Lebensrisiko“ aufbürden, wenn Softwareagenten Ungewissheitsentscheidungen treffen dürfen? Und dies auch noch mit der „humanistischen“ Begründung, dass nur Menschen, nicht Computer im Rechtssinne handeln können?<sup>84</sup>

### III. RECHTSPROBLEME ZUM AUTONOMIERISIKO

Als Zwischenergebnis lässt sich festhalten. Das Autonomierisiko stellt das Privatrecht vor neue Herausforderungen, freilich nicht in dem Sinne, dass es die volle Personifizierung von Softwareagenten nahelegt. Vielmehr sollte den Softwareagenten ein sorgfältig kalibrierter Rechtsstatus zuerkannt werden. Die Antwort auf das Autonomierisiko wäre ihr Status als Aktanten, als teilrechtsfähige Akteure, deren autonome Entscheidungen mit Rechtsverbindlichkeit ausgestattet werden und Haftungsfolgen auslösen. Dieser neue Rechtsstatus soll nun an ausgewählten Einzelproblemen ausgeformt werden. Dabei steht im Vordergrund, ob und wie die eingangs geschilderten Verantwortungsdefizite bewältigt werden können. Aber es wird zugleich darum gehen, wie ein solcher Rechtsstatus von Softwareagenten ermöglicht, die Dogmatik der Rechtsgeschäftslehre und des vertraglichen und außervertraglichen Haftungsrechts in sich stimmig und frei von Fiktionen weiterzuentwickeln.

#### 1. Digitale Verträge

Für die Geschäftspraxis bedeutet es buchstäblich eine revolutionäre Umwälzung, dass Menschen an Algorithmen die Aufgabe delegieren, selbständig Verträge abzuschließen und durchzuführen. Zugleich aber ist das Vertragsrecht in seinen dogmatischen Fundamenten berührt, denn seine selbstverständliche Voraussetzung war bisher, dass ausschließlich Menschen – und das gilt auch für die Rechtsakte juristischer Personen – Willenserklärungen abgeben, besonders Verträge abschließen können. Deshalb war es durchaus konsequent,

responsibility – when software agents are allowed to make uncertain decisions? And justify this with the “humanistic” reasoning that only people, not computers, can act in the legal sense?<sup>84</sup>

### III. LEGAL PROBLEMS CONCERNING THE AUTONOMY RISK

As an interim result, it can be said: The autonomy risk poses new challenges to private law, but not in the sense that it suggests full legal personality for software agents. Rather, software agents should be given a carefully calibrated legal status. The answer to the autonomy risk would be their status as actants, as actors with partial legal personhood, whose autonomous decisions are made legally binding and in case they are unlawful trigger liability consequences. This new legal status will now be shaped on selected legal issues. The focus is on whether and how the above-mentioned deficiencies in responsibility can be overcome. But at the same time, it will be a question of how such a legal status of software agents makes it possible to develop the doctrine of legal transactions and contractual and non-contractual liability coherently and free of fictions.

#### 1. Digital contracts

In business practice, it is literally a revolution when people delegate the task of negotiating and executing contracts to algorithms. At the same time, however, contract law is affected at its foundations, because it has been a matter of course that only human individuals – and this also applies to the legal acts of legal persons – were able to conclude contracts. For this reason, it was quite logical for some authors to insist that autonomous computer declarations are void under current law until the legislation allows for them and provides detailed re-

84 *Stefan Kirn/Claus D. Müller-Hengstenberg*, *Rechtliche Risiken autonomer und vernetzter Systeme: Eine Herausforderung* (2016), 3.3.7 bei Fn. 799 mit weiteren Nachweisen. sind für diese widersprüchliche Argumentation typisch. Wie stattdessen eine Rehumanisierung der technischen Welt – auch mit Mitteln des Rechts – eingeleitet werden könnte, dazu eingehend *Gruber* (Fn. 13), 323 ff.

84 *Stefan Kirn/Claus D. Müller-Hengstenberg*, *Rechtliche Risiken autonomer und vernetzter Systeme: Eine Herausforderung* (2016), 3.3.7 at fn. 799 with further references, are typical of this contradictory argumentation. How instead a re-humanization of the technical world – also by means of law – could be initiated, in detail, *Gruber* (fn. 13), 323 ff.



wenn einige Autoren darauf insistierten, dass autonome Computererklärungen nach geltendem Recht unwirksam sind, solange bis die Gesetzgebung sie zulässt und eine eingehende Regelung erlässt, die Verbindlichkeits- und Haftungsfragen regelt.<sup>85</sup> Genau dies tat die Gesetzgebung in USA und Kanada. Section 14 des Uniform Electronic Transactions Act erklärt Verträge zwischen Algorithmen für wirksam:

„A contract may be formed by the interaction of electronic agents of the parties, even if no individual was aware of or reviewed the electronic agents' actions or the resulting terms and agreements.“<sup>86</sup>

Auch das Europäische Parlament erklärt die überkommenen Regeln des Vertragsrechts für inadäquat und spricht von der „Notwendigkeit für neuere aktuellere Regeln“.<sup>87</sup> Das deutsche Recht hingegen sieht hier keinerlei fundamentale Probleme. Es begnügt sich damit, für Computererklärungen ein paar Verbraucherschützende Formvorschriften vorzusehen. Als Grund wird angegeben, die Rechtsgeschäftslehre verfüge über einen solchen inneren Reichtum, dass auch Computererklärungen „mit tradierten rechtlichen Instrumenten bewältigt werden“ können.<sup>88</sup> Gegen den Protest einzelner Autoren erklärt die herrschende Lehre: Die automatisierte Willenserklärung „ist“ (!) die Erklärung dessen, der die EDV-Anlage betreibt, da die Anlage keinen eigenen Willen betätigt, sondern aufgrund ihrer Programmierung (zumindest mittelbar) den Willen des Anlagenbetreibers „verlautbart“.<sup>89</sup> Der BGH bestätigte dies, wenn auch nur in einem *obiter dictum*: „Nicht das Computersystem, sondern die Person (oder das Unternehmen), die es als Kommunikationsmittel nutzt, gibt die Erklärung ab oder ist Empfänger der abgegebenen Erklärung.“<sup>90</sup>

Aber gilt dies auch im eigentlich kritischen Fall, wenn der Softwareagent nicht voll determiniert ist, sondern selbst autonom über den Vertrag entscheidet, also eigenmächtig bestimmte Angebote sucht, Verhandlungen mit potentiellen Partnern führt, den Vertragspartner bestimmt, über den Vertragsabschluss entscheidet, die *essentialia* der Willenserklärung konkretisiert, weiterhin den Vertragsab-

gulations on bindingness and liability.<sup>85</sup> This is what the legislation in the USA and Canada has done. Section 14 of the Uniform Electronic Transactions Act declares contracts between algorithms legally valid:

“A contract may be formed by the interaction of electronic agents of the parties, even if no individual was aware of or reviewed the electronic agents' actions or the resulting terms and agreements.“<sup>86</sup>

The European Parliament also declares the traditional rules of contract law to be inadequate and speaks of the “need for new, efficient and up-to-date” rules.<sup>87</sup> German law, on the other hand, does not see any fundamental problems here. It is enough to provide for a few consumer-protecting formalities for computer declarations. The reason given is that contract doctrine possesses such an inner richness that computer declarations can also be “handled with traditional legal instruments”.<sup>88</sup> Against the protest of a few authors, the prevailing doctrine asserts: The automated declaration of intent “is” (!) the declaration of the human person who operates the computer system, since the system does not act on its own will, but rather “announces” the will of the operator on the basis of its programming (at least indirectly).<sup>89</sup> The German Federal Supreme Court has confirmed this, albeit only in an *obiter dictum*: “Not the computer system, but the person (or the company) who uses it as a means of communication and thus makes the declaration or is the recipient of the declaration made.”<sup>90</sup>

But does this also apply in the actually critical case, if the software agent is not fully determined but decides autonomously on the contract itself, i.e. searches for offers on his own authority, negotiates with potential partners, chooses the contractual party, decides on the conclusion of the contract, determines the *essentialia* of the transaction, determines the expiry of the contract, exercises with-

85 Für Unwirksamkeit: *Wettig* (Fn. 17), 162 f.; *Rotraut Gitter*, Softwareagenten im elektronischen Rechtsverkehr (2007), 173.

86 Section 14 of the Uniform Electronic Transactions Act, <http://www.law.upenn.edu/bll/ulc/fnact99/1990s/ueta99.pdf>. Sec. 21 of the Canadian Uniform Electronic Commerce Act, <http://www.law.ualberta.ca/alri/ulc/current/euecafin.htm>.

87 Europäisches Parlament (Fn. 3), 7.

88 Münchener Kommentar BGB/Busche, 7. Aufl. 2015, vor § 145, 37.

89 Münchener Kommentar BGB/Busche, 7. Aufl. 2015, vor § 145, 37 und Einl. Rn. 185; *Gerald Spindler/Fabian Schuster*, Recht der elektronischen Medien. Kommentar, 3. Aufl. (2015), vor §§ 116 ff. BGB Rn. 9; *Christian Döpke*, The Importance of Big Data for Jurisprudence and Legal Practice, in: Hoeren/Kolany-Raiser (Hg.), Big Data in Context. Springer Briefs in Law (2018), 1–19, 15.

90 BGHZ 195, 126, Rn. 17.

85 For invalidity, *Wettig* (fn. 17), 162 f.; *Rotraut Gitter*, Softwareagenten im elektronischen Rechtsverkehr (2007), 173.

86 Section 14 of the Uniform Electronic Transactions Act, <http://www.law.upenn.edu/bll/ulc/fnact99/1990s/ueta99.pdf>. Sec. 21 of the Canadian Uniform Electronic Commerce Act, <http://www.law.ualberta.ca/alri/ulc/current/euecafin.htm>.

87 European Parliament (fn. 3), 7.

88 Münchener Kommentar BGB/Busche, 7th ed. 2015, BGB-AT vor § 14, 37.

89 Münchener Kommentar BGB/Busche, 7th ed. 2015, BGB-AT vor § 145, 37 and Einleitung Rn. 185; *Gerald Spindler/Fabian Schuster*, Recht der elektronischen Medien. Kommentar, 3rd ed. (2015), vor §§ 116 ff. BGB Rn. 9; *Christian Döpke*, The Importance of Big Data for Jurisprudence and Legal Practice, in: Hoeren/Kolany-Raiser (ed.), Big Data in Context. Springer Briefs in Law (2018), 1–19, 15.

90 BGHZ 195, 126, Rn. 17.

lauf bestimmt, Rücktritt und andere Gestaltungsrechte ausübt, bei Vertragsbruch Sanktionen festlegt? Und gilt dies auch dann, wenn auf der Gegenseite ein elektronischer Softwareagent existiert? Besonders im Vertragsgeschehen ist das oben allgemein charakterisierte Autonomierisiko sehr viel höher als das bloße Automatisierungsrisiko. Es sind zwar immer noch die Menschen, die über das Computer-Programm und über dessen generellen Einsatz bestimmen. Aber in der darauffolgenden Vielzahl von Verträgen entzieht sich das konkrete Verhalten des Softwareagenten der Kontrolle des Betreibers. Selbst vom Programmierer ist es nicht mehr im einzelnen determinierbar, nicht mehr prognostizierbar und nicht mehr kontrollierbar.<sup>91</sup> Doch selbst für diese Situation eines Kontrollverlusts durch Programmierer/Hersteller/Betreiber hält die herrschende Lehre daran fest, dass ausschließlich der menschliche Geschäftsherr selbst die Erklärung abgegeben hat.<sup>92</sup>

Dies ist eine unhaltbare Fiktion.<sup>93</sup> Und zwar eine Fiktion, die – anders als sonst im Falle von legitimen Rechtsfiktionen – nicht einmal offengelegt wird. Stattdessen wird es als unbestreitbare Tatsache ausgegeben, als „alternativer Fakt“ sozusagen, dass nur der Geschäftsherr selbst die Erklärungen abgibt. Damit verzichtet die herrschende Lehre auf eine überzeugende dogmatische Lösung. Sie schwankt dabei zwischen verschiedenen wenig plausiblen Konstruktionsversuchen.

Wie ist es möglich, dass eine Vertragsrechtsdoktrin, die in der Vergangenheit erfolgreich auf die Herausforderungen des modernen entpersonalisierten Geschäftsverkehrs mit einer differenzierten Theorie objektivierter Erklärungen reagiert hat, auf die digitale Herausforderung nur noch defensiv mit uneingestandenen Fiktionen antwortet? Welch ein Widerspruch, dass das deutsche Privatrecht auf der einen Seite klammheimlich Verträge zwischen Maschinen als rechtsgültig anerkannt hat, auf der anderen Seite aber nicht bereit ist, aus der Delegation von Vertragsentscheidungen an Softwareagenten die notwendigen Konsequenzen für tragende Elemente des Vertragsrechts zu ziehen. Man beschwört Willensfreiheit und Menschenwürde, um jede Art von Rechtssub-

drawal, lays down sanctions in the event of breach of contract? And does this also apply if there is an electronic software agent on the opposite side? Particularly in contracts, the autonomy risk, which was characterized generally above, is much higher than mere automation risk. To be sure, it is still the human actors who decide about the computer program and its general use. But in the multitude of contracts that follow, the concrete behavior of the software agent is beyond the control of the operator. Even for the programmer it is no longer determinable in detail, no longer predictable and no longer controllable.<sup>91</sup> But even in this situation of a loss of control by programmer/manufacturer/operator, the prevailing doctrine maintains that it is only the principal himself who has made the declaration.<sup>92</sup>

This is an untenable fiction.<sup>93</sup> This is a fiction that – unlike in the case of legitimate legal fictions – is not even revealed. Instead, it is presented as an indisputable fact, as an “alternative fact”, so to speak, that only the principal himself makes the declaration. Thus, the prevailing doctrine dispenses with a convincing legal solution. In doing so, it fluctuates between various construction attempts that are rather implausible.

How is it possible that a contract law doctrine, which in the past successfully responded to the challenges of modern depersonalized business with a sophisticated theory of objective declarations, can only respond defensively to the digital challenge via unrevealed fictions? What a contradiction that German private law has tacitly recognized contracts between machines as legally valid, but on the other hand is not prepared to draw the consequences for contract law that follow from the delegation of contractual decisions to software agents. Instead, freedom of will and human dignity are invoked to block any kind of legal subjectivity for software agents as unconstitutional (!): “The fundamentals of the constitution cannot justify any recognition of a legal personality with legal capacity for

91 *Matthias* (Fn. 4), 33 ff.

92 Kommentarliteratur siehe Fn. 89. Ebenso *Kai Cornelius*, Vertragsabschluss durch autonome elektronische Agenten, *Multimedia und Recht* 5 (2002), 353–358, 355; *Kirn/Müller-Hengstenberg* (Fn. 2), 67. Das soll auch explizit dann gelten, wenn der Softwareagent die *essentialia* des Geschäfts konkretisiert, *Susanne Horner/Marcus Kaulartz*, Haftung 4.0: Rechtliche Herausforderungen im Kontext der Industrie 4.0, *InTeR Zeitschrift zum Innovations- und Technikrecht* (2016), 22–27, 22.

93 Zur Kritik der Fiktion *Francisco Andrade/Paulo Novais/José Machado/José Neves*, Contracting Agents: Legal Personality and Representation, *Artificial Intelligence Law* 15 (2007), 357–373, 360; *Allen/Widdison* (Fn. 17), 46 ff.

91 *Matthias* (fn. 4), 33 ff.

92 Commentary literature see fn. 89. Also *Kai Cornelius*, Vertragsabschluss durch autonome elektronische Agenten, *Multimedia und Recht* 5 (2002), 353–358, 355; *Kirn/Müller-Hengstenberg* (fn. 2), 67. This shall also apply explicitly if the software agent concretizes the *essentialia* of the contract, *Susanne Horner/Marcus Kaulartz*, Haftung 4.0: Rechtliche Herausforderungen im Kontext der Industrie 4.0, *InTeR Zeitschrift zum Innovations- und Technikrecht* (2016), 22–27, 22.

93 For a critique of the fiction, *Francisco Andrade/Paulo Novais/José Machado/José Neves*, Contracting Agents: Legal Personality and Representation, *Artificial Intelligence Law* 15 (2007), 357–373, 360; *Allen/Widdison* (fn. 17), 46 ff.

jektivität für Softwareagenten als verfassungswidrig (!) zu blockieren: Es könne das „Grundbekenntnis der Verfassung keine Anerkennung einer Rechtspersönlichkeit mit Rechtsfähigkeit für diese Softwareagenten rechtfertigen.“<sup>94</sup> Zugleich aber – und darin steckt die Ironie – werden angesichts der faktischen Delegation wesentlicher Entscheidungen an seelenlose Maschinen die hohen ethischen Ansprüche an Privatautonomie, wonach menschliche Subjekte das Vertragsgeschehen eigenverantwortlich beherrschen sollen, gerade mit Hilfe dieser Fiktion schlicht aufgegeben.

Deshalb sollte man konsequent auf die Gegenposition übersetzen, wie es eine Reihe von Autoren vertreten:<sup>95</sup> In genauer Entsprechung zu seiner realen Funktion in der Wirtschaftspraxis, ökonomisch gesprochen, in einer *Principal-agent-Beziehung*<sup>96</sup>, gibt der Softwareagent rechtlich eigenverantwortlich die Willenserklärung ab, handelt aber nicht im eigenen Namen, sondern im Namen des Prinzipals. Die Erklärung des Computers ist als solche rechtsverbindlich und berechtigt und verpflichtet unmittelbar den Prinzipal. In Analogie zum Recht der Stellvertretung sind autonome Softwareagenten als Repräsentanten ihres menschlichen Prinzipals zu behandeln. Dafür braucht man ihnen – wie schon vorher gesagt – keine volle Rechtsfähigkeit als juristische Person zuzuschreiben, sondern in funktionaler Sicht genügt die bloße Teilrechtsfähigkeit, die Stellvertretungsfähigkeit.<sup>97</sup> Dass die Verleihung der Rechtsfähigkeit schon *de lege lata* möglich ist, zeigt die Geschichte des nicht-rechtsfähigen Vereins, besonders der Gewerkschaften, der Vorgesellschaft und jüngst die Entscheidungen zur Rechtsfähigkeit der BGB-Gesellschaft. Dies sind Präjudizien, in denen die Gerichte *praeter legem*, wenn nicht *extra legem*, (beschränkte) Rechtsfähigkeit an zuvor nicht-rechtsfähige Entitäten verliehen haben.<sup>98</sup>

Mit Hilfe einer solchen Analogie lässt sich der wohl wichtigste Einwand ausräumen, dass den Softwareagenten die notwendigen Willenselemente fehlen.<sup>99</sup> Mit gebotener Sorgfalt ist es durchaus möglich, digitale Äquivalente für die sogenannten

these software agents.“<sup>94</sup> At the same time, however – and this is the irony – in view of the factual delegation of essential decisions to soulless machines, the profound ethical demands of private autonomy, according to which human subjects are supposed to master the contractual process on their own authority, are simply abandoned precisely with the help of this fiction.

Therefore, one should consistently move to the opposite position, as a number of authors argue:<sup>95</sup> Corresponding exactly to its real function in business practice, economically speaking, in a *principal-agent relationship*<sup>96</sup>, the software agent concludes the contract on its legal authority, but does not act in its own name, but on behalf of the principal. The declaration of the computer as such is legally binding while authorizing and directly obligating the principal. In analogy to the law of agency, autonomous software agents are to be treated as representatives of their human principal. In order to do so, as has already been said, they do not have to be ascribed full legal personhood as a legal entity, but, from a functional point of view, mere partial legal capacity, i.e. the ability to act as a representative, suffices.<sup>97</sup> Granting legal personhood is already possible *de lege lata*, this shows the legal history of associations originally without legal status, especially the trade unions, the companies in formation and most recently the decisions on the legal capacity of partnerships. These are precedents in which the courts have conferred *praeter legem*, if not *extra legem*, (limited) legal capacity on entities which previously lacked legal status.<sup>98</sup>

With the help of such an analogy, the arguably most important objection that software agents lack the necessary contractual intention can be dispelled.<sup>99</sup> It is quite possible to find digital equivalents for the subjective preconditions of the declaration of

94 *Kirn/Müller-Hengstenberg* (Fn. 84), 2.2.1 bei Fn. 351.

95 *Beck* (Fn. 3), 186; *Schirmer* (Fn. 23), 664; *Kersten* (Fn. 17), 7; *Gruber* (Fn. 28, 2015), 198; *Gruber* (Fn. 28, 2012), 140 ff.; *Wettig* (Fn. 17), 179 ff.; *Gitter* (Fn. 85), 177 ff.; *Robert John*, Haftung für künstliche Intelligenz (2007), 77 ff., 154 f.; *Teubner* (Fn. 29); *Sorge* (Fn. 2), 118; *Georg Schwarz*, Die rechtsgeschäftliche ‚Vertretung‘ durch Softwareagenten, in: Schweighofer (Hg.), Auf dem Weg zur ePerson, 67 (2001), 67 f.; *Fischer*, Computers as Agents: A Proposed Approach to Revised U.C.C. Article 2, *Indiana Law Journal* 72 (1997), 557. Für Stellvertretung, aber nur *de lege ferenda* *Oliver Kefler*, Intelligente Roboter – neue Technologien im Einsatz: Voraussetzungen und Rechtsfolgen des Handelns informationstechnischer Systeme, *Multimedia und Recht* (2017), 589–594, 592.

96 Zur ökonomischen Definition der *Principal-Agent-Beziehung*, die weitgehend der rechtlichen Qualifikation als Stellvertretung entspricht: *Jensen/Meckling* (Fn. 24).

97 Zur durchaus vergleichbaren Konstruktion der Rechtsgeschäfte von Sklaven im römischen Recht *Harke* (Fn. 25), 97 f.

98 Zu dieser Parallele eingehend *Gruber* (Fn. 13), 267 ff. Vgl. BGHZ 146, 341, 344.

99 Zu diesem Einwand *Spindler* (Fn. 12), 816.

94 *Kirn/Müller-Hengstenberg* (fn. 84), 2.2.1, fn. 351.

95 *Beck* (fn. 3), 186; *Schirmer* (fn. 23), 664; *Kersten* (fn. 17), 7; *Gruber* (fn. 28, 2015), 198; *Gruber* (fn. 28, 2012), 140 ff.; *Wettig* (fn. 17), 179 ff.; *Gitter* (fn. 85), 177 ff.; *Robert John*, Haftung für künstliche Intelligenz (2007), 77 ff., 154 f.; 154 f.; *Teubner* (fn. 29); *Sorge* (fn. 2), 118; *Georg Schwarz*, Die rechtsgeschäftliche ‚Vertretung‘ durch Softwareagenten, in: Schweighofer (ed.), Auf dem Weg zur ePerson, 67 (2001), 67 f.; *Fischer* (1997) „Computers as Agents: A Proposed Approach to Revised U.C.C. Article 2“, 72 *Indiana Law Journal*, 557. For representation, but only *de lege ferenda* *Oliver Kefler*, Intelligente Roboter – neue Technologien im Einsatz: Voraussetzungen und Rechtsfolgen des Handelns informationstechnischer Systeme, *Multimedia und Recht* (2017), 589–594, 592.

96 On the economic definition of the principal-agent relationship, which largely corresponds to the legal qualification in agency law, *Jensen/Meckling* (fn. 24).

97 On the comparable construction of the legal transactions of slaves in Roman law, *Harke* (fn. 25), 97 f.

98 Detailed to this parallel, *Gruber* (fn. 13), 267 ff. See BGHZ 146, 341, 344.

99 For this objection, e.g., *Spindler* (fn. 12), 816.

subjektiven Voraussetzungen der Willenserklärung, besonders für das Erklärungsbewusstsein, zu finden. An dieser Stelle kommen die bekannten Objektivierungstendenzen in der Rechtsgeschäftslehre der Möglichkeit, dass bewusstseinslose Softwareagenten rechtswirksame Willenserklärungen abgeben können, weit entgegen.<sup>100</sup> Sie bestätigen die schon angesprochene allgemeine soziologische These, dass es für die Personifizierung nicht-menschlicher Entitäten nicht auf Bewusstsein, sondern auf Kommunikation ankommt. Bekanntlich hat die Auflösung der Kontroverse Willens- theorie versus Erklärungstheorie in modernen Geltungstheorien und Vertrauens- theorien das Erklärungsbewusstsein entpsychologisiert.<sup>101</sup> Auf den subjektiven Willen des Erklärenden kommt es nach der dramatischen Wende in der neueren Rechtsprechung gar nicht mehr an. Der objektive Vertrauensgrundsatz hat die subjektiven Vorstellungen der Parteien verdrängt. Entscheidend ist, dass sich der Erklärende sein falsches äußeres Verhalten im Rechtsverkehr zurechnen lassen muss. Der Erklärende kann sein fehlendes Erklärungsbewusstsein nicht der Geltung der Willenserklärung entgegensetzen, wenn er, wie der BGH entschieden hat, „fahrlässig nicht erkannt hat, dass sein Verhalten als Willenserklärung aufgefasst werden könnte und wenn der Empfänger es tatsächlich so verstanden hat“.<sup>102</sup>

An die Stelle des subjektiven Erklärungsbewusstseins setzt also der BGH eine doppelte objektive Normierung. Damit aber können Softwareagenten, die über elaborierte kognitive Fähigkeiten verfügen, umgehen. Die Normierung besteht aus zwei Normen: erstens, die aus dem Vertrauen entstandene soziale Norm, ob das konkrete Verhalten als eine bindende Willenserklärung verstanden werden darf, zweitens, die Pflicht des Erklärenden, diese soziale Norm zu erkennen und nicht gegen diese Pflicht zur Kenntnisnahme fahrlässig zu verstoßen. Eine solche Kenntnis sozialer Normen, also wie in einem bestimmten Kontext bestimmte Erklärungen zu verstehen sind, kann durchaus in ein Softwareprogramm übersetzt werden. Ja, lernfähige Agenten können diese Information sogar selbsttätig erwerben und sie darüber hinaus bei Änderungen der sozialen Normen oder der Rechtsprechung selbsttätig abändern. Damit ist der Einwand, Softwareagenten als digitale Stellvertreter könnten kein Erklärungsbewusstsein besitzen, weil sie kein Bewusstsein haben, ausgeräumt. Das Äquivalent sind ihre kognitiven Fähigkeiten.

intent. At this point, the well-known objectivization tendencies in contract theory meet the possibility that software agents without consciousness of their own can make legally effective declarations of intent.<sup>100</sup> They confirm the general sociological thesis mentioned above, that it is not consciousness, which is important for the personification of non-human entities, but communication. It is well known that modern reliance theories have replaced the controversy of will theory versus declaration theory and have de-psychologized the contractual intent.<sup>101</sup> After the dramatic change in recent court decisions, the subjective will of the declaring party no longer matters at all. The objective principle of reliance has replaced the subjective intentions of the parties. The deciding criterion is that the external behavior can be attributed to the contracting person. The declaring party cannot invoke his lacking intention to the validity of the declaration if he has, as the Federal Supreme Court has decided, “negligently failed to recognize that his behavior could be understood as a declaration of intent and if the recipient has actually understood it so”.<sup>102</sup>

The BGH thus places two objective standards instead of a subjective intention. Indeed, software agents with elaborate cognitive abilities can handle this. The standardization consists of two norms: firstly, the social norm based on trust, whether the concrete behavior may be understood as a binding declaration of intent; and secondly, the obligation of the declaring party to recognize this social norm and not to negligently violate this obligation to acknowledge it. Such knowledge of social norms, i.e. how to understand certain declarations in a particular context, can be translated into a software program. Moreover self-learning agents can even acquire this information on their own initiative and modify it themselves in the event of changes in social norms or case law. Thus, the objection that software agents as digital representatives could not possess contractual intent because they have no consciousness has been dispelled. The equivalent is their cognitive abilities.

100 *Kirn/Müller-Hengstenberg* (Fn. 84), 2.2.

101 Z. B. Münchener Kommentar BGB/Säcker 7. Auflage 2015 BGB-AT Einleitung, Rn. 158–162 mit weiteren Nachweisen.

102 BGH NJW 1995, 953.

100 *Kirn/Müller-Hengstenberg* (fn. 84), 2.2.

101 E.g., Münchener Kommentar BGB/Säcker 7th ed. 2015 BGB-AT Einleitung, Rn. 158–162 with further references.

102 BGH NJW 1995, 953.

## 2. Vertragliche Haftung

Bei den eben angesprochenen Computer-Erklärungen ging es um die Verbindlichkeit von Verträgen, die Softwareagenten für einen Prinzipal abschließen. Nun geht es um die Situation, in der Menschen oder Unternehmen Verträge abschließen und zur Vertragserfüllung Softwareagenten heranziehen. Wie ist dann eine Schadensverursachung durch den Computer zu beurteilen? Verstößt das Verhalten autonomer Softwareagenten gegen Vertragspflichten, insbesondere schädigt es die Rechtsgüter des Vertragspartners, dann gerät die Lehre, die sich weigert, den Agenten die beschränkten Handlungsfähigkeiten eines Erfüllungsgehilfen zuzusprechen, in große Schwierigkeiten.

Da die Lehre an der Prämisse unbeirrt festhält, dass selbst autonome Softwareagenten nicht handeln können und deshalb auch keine Erfüllungsgehilfen sein können,<sup>103</sup> kann sie die Vertragspflichtverletzung nicht am eigentlich schädigenden Verhalten des Softwareagenten festmachen, sondern muss für sämtliche Einzelelemente der Vertragsverletzung auf die Pflichtverletzung durch den Betreiber selbst abstellen.<sup>104</sup> Dies aber ist eine dogmatische Fehlkonstruktion, denn als schuldhaftige Vertragsverletzung kann dann dem Betreiber, wie Kritiker zu Recht monieren, nur die Inbetriebnahme (!) des Computers vorgeworfen werden.<sup>105</sup> Gegenüber einer solchen verfehlten Anknüpfung an die Inbetriebnahme als Pflichtverletzung bestehen nicht nur rechtsdogmatische Bedenken. Rechtspolitisch wird dagegen eingewandt, dass damit jede Steuerungszweckhaftigkeit der Verschuldenshaftung leerläuft. Zudem sei es innovationsfeindlich, wenn der bloße Betrieb zum Haftungsgrund gemacht wird.<sup>106</sup>

Schlimmer aber ist: Bei dieser dogmatisch und rechtspolitisch fragwürdigen Konstruktion entsteht zugleich eine schwer erträgliche Haftungslücke, die sich in Zukunft noch ausweiten wird, je mehr Aufgaben der Vertragserfüllung an autonome Softwareagenten delegiert werden.<sup>107</sup> „Kann der Betreiber aber nachweisen, dass der Schadenseintritt nach dem Stand der Technik weder vorhersehbar noch vermeidbar war, so entfällt [...] eine Haftung.“<sup>108</sup>

## 2. Contractual liability

The computer declarations just mentioned concerned the binding nature of contracts concluded by software agents for a principal. Now we turn to the situation in which people or companies conclude contracts and use software agents as vicarious agents to fulfill the contractual obligations. How can a damage caused by the computer be legally sanctioned? If the behavior of the autonomous software agent violates contractual obligations, in particular if it damages the assets of the contractual partner, then the doctrine that refuses to assign the limited capacity of a vicarious agent to the software agent becomes very difficult.

Since the doctrine unwaveringly adheres to the premise that even autonomous software agents have no legal action capacity and therefore cannot be vicarious agents,<sup>103</sup> the breach of contractual obligations cannot be attributed to the actually damaging behavior of the software agent, but must instead be found in all its individual elements for a breach of duty in the behavior of the human actor involved.<sup>104</sup> But this is a dogmatic misconception, because as a breach of contract then the operator, as critics rightly bemoan, can only be accused of putting the computer into operation (!).<sup>105</sup> Beyond legal doctrine it can be objected from a policy perspective that this excludes any prevention intention of the fault-based liability. In addition, it is hostile to innovation if mere putting into operation is made the cause of liability.<sup>106</sup>

Worse, however, is: With this construction a liability gap arises that is difficult to bear and which will widen in the future as more tasks of contract performance are delegated to autonomous software agents.<sup>107</sup> “If the operator can prove, however, that the occurrence of damage was neither predictable nor avoidable in accordance with the state of the art, then [...] liability is omitted.”<sup>108</sup>

103 *Palandt/Grüneberg*, BGB 77. Aufl. 2018, § 278 Rn. 11; *Helmut Redeker*, IT-Recht (2012), Rn. 969 f.; *Koch*, Internet-Recht (2005), § 3, Nr. 1 b, S. 101; *Dirk Heckmann*, Internetrecht (2007), Kap. 4.1 Rn. 18.

104 *Horner/Kaulartz* (Fn. 92), 23; *Jochen Hanisch*, Zivilrechtliche Haftungskonzepte für Robotik, in: Hilgendorf (Hg.), Robotik im Kontext von Recht und Moral (2014), 27–61, 32.

105 Dies kritisieren auch *Lohmann* (Fn. 49), 158; *Schirmer* (Fn. 23), 664.

106 *Hanisch* (Fn. 104), 34.

107 Auch Münchener Kommentar BGB/Grundmann, 7. Aufl. 2016, BGB § 278 Rn. 46 spricht von einer Regelungslücke, die durch die technologische Entwicklung bedingt ist.

108 *Kirn/Müller-Hengstenberg* (Fn. 9), 16. Ebenso *Palandt/Grüneberg* BGB 77. Aufl. 2018, § 276 Rn. 20 f.

103 *Palandt/Grüneberg*, BGB 77th ed. 2018, § 278 Rn. 11; *Helmut Redeker*, IT-Recht (2012), Rn. 969 f.; *Koch*, Internet-Recht (2005), § 3, Nr. 1 b, S. 101; *Dirk Heckmann*, Internetrecht (2007), Kap. 4.1 Rn. 18.

104 *Horner/Kaulartz* (fn. 92), 23; *Jochen Hanisch*, Zivilrechtliche Haftungskonzepte für Robotik, in: Hilgendorf (ed.), Robotik im Kontext von Recht und Moral (2014), 27–61, 32.

105 This is also criticized by *Lohmann* (fn. 49), 158; *Schirmer* (fn. 23), 664.

106 *Hanisch* (fn. 104), 34.

107 Münchener Kommentar BGB/Grundmann, 7<sup>th</sup> ed. 2016, BGB § 278 Rn. 46 speaks of a regulatory gap caused by technological developments.

108 *Kirn/Müller-Hengstenberg* (fn. 9), 16. Also *Palandt/Grüneberg* BGB 77th ed. 2018, § 276 Rn. 20 f.

Beide Schwierigkeiten – die Haftungslücke ebenso wie die dogmatische Fehlkonstruktion – lassen sich dagegen problemlos vermeiden, wenn man die Analogie wagt und die Gehilfenhaftungsregel analog auf das Verhalten autonomer Softwareagenten anwendet. Auch wenn dem Geschäftsherrn selbst kein Fehlverhalten vorzuwerfen ist, haftet er dann für jedes Verhalten des Softwareagenten, sofern dieses die Vertragspflichten verletzt und einen Schaden verursacht.<sup>109</sup> Die Analogie würde, wie Grundmann aus ökonomischer Sicht darlegt, positiv zur Internalisierung der Kosten von Maschinenversagen beitragen, und da der Entscheidungszugleich Kostenträger wäre, hätte er den optimalen Anreiz, Nutzen und Kosten von mehr Maschinensicherheit minimierend abzuwägen.<sup>110</sup>

Die Analogie beseitigt die Haftungslücke vollständig. Denn der Geschäftsherr kann sich nicht mehr damit entlasten, ihm selbst sei kein Fehlverhalten vorzuwerfen. Es kommt allein auf das Fehlverhalten des Softwareagenten an.<sup>111</sup> Hier liegt der eigentliche Vorzug der Analogie zur Gehilfenhaftung gegenüber der Haftung des Prinzipals für sein eigenes Fehlverhalten. Denn auch dann, wenn der Prinzipal alle Sorgfaltspflichten beim Einsatz des Computers erfüllt hat, muss er dennoch für ein Entscheidungsversagen des autonomen Softwareagenten haften, genauso wie wenn ein menschlicher Erfüllungshelfer gehandelt hätte.

Der tiefere Grund liegt, wie gesagt, darin, dass die Gesellschaft, wenn sie neuartige Entscheidungsräume für bis dahin unbekannte autonome Entscheidungsträger zulässt, verpflichtet ist, für wirksame Verantwortungsformen zu sorgen. Entscheidend sind letztlich nicht Gesichtspunkte von Effizienz, von Transaktionskostensparnissen, von utilitaristischen Erwägungen, von soziologischer Jurisprudenz oder von *policy-questions*, sondern dies ist eine genuine Frage juridischer Gerechtigkeit. Es ist der Gleichbehandlungsgrundsatz, der hier die Haftung verlangt. Denn würde ein Mensch für die Vertragsdurchführung herangezogen, so haftete der Prinzipal für dessen Pflichtverletzungen, er kann aber nicht von der Haftung befreit sein, wenn für die identische Aufgabe ein Softwareagent herangezogen wird. Diese nicht zu rechtfertigende Privilegierung, die darin besteht, dass die Digitalisierung einem Computerbetreiber einen solchen erheblichen Kostenvorteil zulasten der geschädigten Vertragspartner verschafft, müsste man der Rechtsdogmatik, die unbeirrt an tradierten Rechtsbegriffen festhält, vorwerfen.

109 Für die Analogie *Keßler* (Fn. 95), 592 f.; *Schirmer* (Fn. 23), 665; *Gruber* (Fn. 28, 2015), 198; *Teubner* (Fn. 29); vorsichtiger nur für die Zukunft befürwortend *Günther* (Fn. 17), 84.

110 Münchener Kommentar BGB/Grundmann, 7. Aufl. 2016, BGB § 278 Rn. 46, der sich dann aber doch gegen die Analogie ausspricht.

111 *Schirmer* (Fn. 23), 665.

Both difficulties – the liability gap and the doctrinal misconception – can be avoided without any problems if one applies the liability rules for auxiliary persons analogously to the behavior of autonomous software agents. Even if the principal himself is not accused of misconduct, he shall be liable for any conduct of the software agent, insofar as the latter violates the contractual obligations and causes damage.<sup>109</sup> The analogy would, as Grundmann explains from an economic point of view, internalize the costs of machine failure. Since the decision-maker would also be the cost-bearer, he would have the optimum incentive to weigh up the benefits and costs of greater machine safety in a minimizing manner.<sup>110</sup>

The analogy completely eliminates the liability gap because the principal can no longer relieve himself by alleging lack of misconduct on his part. It depends solely on the misconduct of the software agent.<sup>111</sup> Here lies the real advantage of the liability for auxiliary persons over the principal's liability for his own misconduct. Even if the principal has fulfilled all due diligence obligations when using the computer, he must nevertheless be liable for a decision failure of the autonomous software agent, as if a human vicarious agent had acted.

The deeper reason for this, as I have already said, is that, if society allows new areas of decision-making for previously unknown autonomous decision-makers, it is obliged to ensure effective forms of responsibility. Ultimately, the decisive factors are not aspects of efficiency, transaction cost savings, utilitarian considerations, issues of sociological jurisprudence or policy-questions, but this is a genuine question of legal justice. It is the principle of equal treatment that demands liability here. If the execution of the contract is delegated to a human actor, the principal is liable for his breach of duty, consequently, he cannot be exempted from liability if a software agent is used for the identical task. It would be an unjustifiable privilege, if digitalization provided a computer operator with such a considerable cost advantage to the detriment of the injured contract partner. If legal doctrine unswervingly adheres to traditional legal terms and refuses to assign legal action capacity to software agents, it would have to be accused of treating equal cases unequally.

109 For the analogy, *Keßler* (fn. 29), 592 f.; *Schirmer* (fn. 23), 665; *Gruber* (fn. 28, 2015), 198; *Teubner* (fn. 29); more cautiously advocating only for the future *Günther* (fn. 17), 84.

110 Münchener Kommentar BGB/Grundmann, 7<sup>th</sup> ed. 2016, BGB § 278 Rn. 46, which then, however, opposes the analogy.

111 *Schirmer* (fn. 23), 665.

Subjektive Voraussetzungen sollten bei der Analogie zur Gehilfenhaftung noch weniger Schwierigkeiten machen als bei der Stellvertretung. Für die Gehilfenhaftung kommt es erst recht nicht auf innere Fähigkeiten an, sondern im Prinzip nur auf Kommunikationsteilnahme. Das Problem wird auch hier wieder durch Objektivierungstendenzen im heutigen Privatrecht, welche die Anforderungen an das Verschulden verändert haben, entschärft. Diese Tendenzen lassen in der „objektiven Fahrlässigkeit“ Rechtswidrigkeit und Schuld beinahe miteinander verschmelzen.<sup>112</sup> Das erleichtert die Anforderungen an die Analogie. Denn in der Informationsphilosophie, welche die moralische Verantwortlichkeit von Algorithmen untersucht, argumentiert man, dass die kognitiven Fähigkeiten autonomer Softwareagenten in der Weise entwickelt sind, dass ihnen zwar keine „responsibility“ im Vollsinn zugesprochen werden kann, wohl aber ihre „accountability“.<sup>113</sup> In juristischer Übersetzung hieße dies, dass ihnen Rechtswidrigkeit zugerechnet werden kann, aber kein Verschulden im Sinne von subjektiver Vorwerfbarkeit. Und an dieser Stelle können an Softwareagenten sogar höhere Sorgfaltsanforderungen als an handelnde Menschen gestellt werden, sofern sie wegen ihrer überlegenen Informationsverarbeitungskapazität über höhere kognitive Fähigkeiten verfügen.<sup>114</sup>

### 3. Deliktische Haftung

Im Bereich außervertraglicher Haftung wird das Versagen des geltenden Rechtes gegenüber dem digitalen Autonomierisiko überdeutlich. In aller Härte wird kritisiert, dass hier eine erhebliche Haftungslücke aufgerissen ist, es wird nachdrücklich das Eingreifen der Legislative gefordert.<sup>115</sup> Wegen seiner Bedeutung für die wichtigsten Fälle von Computerversagen ist dies der Testfall dafür, ob man den Softwareagenten partielle Rechtsfähigkeit geben soll oder nicht. Nicht nur, weil die Schadenshäufigkeit hoch und die Schadenshöhe beträchtlich ist. Sondern im Gegensatz zu vertraglicher Haftung, bei denen der Vertragspartner sich immerhin „freiwillig“ dem Risiko aussetzt, einen digitalen Vertragspartner zu haben, ist in Fällen außervertraglicher Haftung der Geschädigte unfreiwillig dem Computerrisiko ausgesetzt, trägt also stellvertretend das Risiko der Gesellschaft.

Subjective intentions should cause even fewer difficulties in the analogy with the liability for auxiliary persons than in the case of contractual representation. For the liability for assistants, it is even less important to have inner abilities than to participate in communication. Here again, the problem is alleviated by objectivization tendencies in today's private law, which have changed the requirements on fault. These tendencies almost merge illegality and fault in “objective negligence”.<sup>112</sup> This simplifies the requirements for the analogy. In information philosophy, which examines the moral responsibility of algorithms, it is argued that the cognitive abilities of autonomous software agents are developed in such a way that they cannot be considered “responsible” in the full sense of the word, but “accountability” can be attributed to them.<sup>113</sup> In legal translation, this would mean that they can be attributed illegality, but no fault in the sense of subjective reproach. And at this point, software agents can even be required to exercise greater care than the acting humans, provided that they possess higher cognitive abilities due to their superior information processing capacity.<sup>114</sup>

### 3. Tortious liability

In the area of non-contractual liability, the failure of the current law to deal with the digital autonomy risk becomes abundantly clear. Authors criticize in all harshness that here too a considerable liability gap has been opened, and they demand urgently legislative intervention.<sup>115</sup> Since most computer failures occur in the field of tortious liability this is the test case for whether or not to give software agents partial personhood. Not only because the frequency of damage is high and the amount of damage is considerable. In contrast to contractual liability, in which the contracting party exposes itself “voluntarily” to the risk of having a digital contractual partner, in cases of non-contractual liability the injured party is involuntarily exposed to the computer risk.

112 Münchener Kommentar BGB/Grundmann, 7. Aufl. 2016, BGB § 276 Rn. 53 ff.

113 Floridi/Sanders (Fn. 6), 205 ff.

114 Münchener Kommentar BGB/Grundmann, 7. Aufl. 2016, BGB § 278 Rn. 50.

115 Europäisches Parlament (Fn. 3), 6.

112 Münchener Kommentar BGB/Grundmann, 7th ed. 2016, BGB § 276 Rn. 53 ff.

113 Floridi/Sanders (fn. 6), 205 ff.

114 Münchener Kommentar BGB/Grundmann, 7<sup>th</sup> ed. 2016, BGB § 278 Rn. 50.

115 European Parliament (fn. 3), 7.

Der eigentliche Grund für das Versagen ist auch hier, dass sich die haftungsrechtlichen Normen ausschließlich darauf konzentrieren, ob Betreibern/Herstellern/Programmierern ein Fehlverhalten vorzuwerfen ist. Dagegen können nach geltendem Recht die Fehlentscheidungen, die Softwareagenten selbst trotz korrekten Verhaltens der menschlichen Beteiligten fällen, nicht sanktioniert werden. Die Haftungslücke entsteht sowohl bei der deliktischen Produzentenhaftung als auch bei der Produkthaftung nach dem Produkthaftungsgesetz, wenn auch in je unterschiedlichem Maße. Der Betreiber ist von der Haftung befreit, wenn er seine Sicherungsvorkehrungen stets an den neuen Stand von Wissenschaft und Technik angepasst hat.<sup>116</sup> Der Produzent haftet nicht, wenn er sämtliche Konstruktionspflichten, Informationspflichten und Produktbeobachtungspflichten erfüllt hat.<sup>117</sup> Besonders fragwürdig erscheint die Haftungsfreistellung für Entwicklungsrisiken angesichts der programmierten (!) Nichtprognostizierbarkeit (!) der Algorithmus-Entscheidungen.

Wenn also die menschlichen Beteiligten ihre Sorgfaltspflichten erfüllt haben, dennoch aber der Softwareagent eine rechtswidrige und schadensverursachende Entscheidung trifft, entsteht die vielkritisierete Haftungslücke. Sie wird in Zukunft wachsen, weil mit zunehmender Verselbständigung autonomer Systeme nicht mehr erwartet werden kann, dass die Nutzer das Verhalten durchgängig überprüfen und entsprechend die Standards der Prüfpflichten sinken werden.<sup>118</sup>

Hier *de lege lata* dennoch eine Haftung zu ermöglichen, dürfte sehr viel schwieriger sein als im Bereich vertraglicher Haftung. Eine denkbare Lösung wäre eine generelle Unternehmenshaftung für die Handlungen autonomer Softwareagenten, wonach deren Fehlverhalten unmittelbar dem Unternehmen zugerechnet wird.<sup>119</sup> Das wäre im Grunde eine Verbundhaftung, die im umfassenden Rahmen der Unternehmensorganisation dem hybriden Charakter der Mensch-Computer-Interaktion mit einem Mitgliedsstatus des Softwareagenten im Verbund gerecht würde. Aber es wäre unrealistisch, dies beim heutigen Stand des Unternehmenshaftungsrechts *de lege lata* vorzuschlagen.

In tort law as well, the liability standards focus exclusively on whether operators/manufacturers/programmers are accused of misconduct. As a consequence, current law does not sanction the wrong decisions made by software agents when the human participants have behaved correctly. The operator is exempted from liability if he has always adapted his safety precautions to the new state of the art in science and technology.<sup>116</sup> The producer is not liable if he has fulfilled all construction, information and product observation obligations.<sup>117</sup> The exemption from liability for development risks seems particularly questionable in view of the programmed (!) non-predictability (!) of the algorithm decisions.

If the human participants have fulfilled their duties of care, but the software agent makes an illegal and damaging decision, the often-criticized liability gap arises. It will grow in the future, because with the increasing independence of autonomous systems it can no longer be expected that users will be able to continuously monitor behavior and the legal standards of testing obligations will decrease accordingly.<sup>118</sup>

However, making liability possible *de lege lata* is likely to be much more difficult than in the area of contractual liability. A conceivable solution would be a general corporate liability for the actions of autonomous software agents, according to which their misconduct is directly attributed to the company.<sup>119</sup> This would be a joint liability, which, within the comprehensive framework of the corporate organization, would do justice to the hybrid character of human-computer interaction with a member status of the software agent in the group. But it would be unrealistic to propose this at the current stage of corporate liability law *de lege lata*.

116 Döpke (Fn. 89), 17; Bräutigam/Klindt (Fn. 12), 1140; Kirn/Müller-Hengstenberg (Fn. 2), 68; Horner/Kaulartz (Fn. 92), 24.

117 Gerald Spindler, Zivilrechtliche Fragen beim Einsatz von Robotern, in: Hilgendorf (Hg.), Robotik im Kontext von Recht und Moral (2014), 63–80, 72 ff., 78.

118 Horner/Kaulartz (Fn. 92), 25.

119 Zu einer überzeugenden rechtspolitischen Konzeption der Unternehmenshaftung Münchener Kommentar BGB/Wagner, 7. Aufl. 2017, BGB, § 831, Rn. 3 f.

116 Döpke (fn. 89), 17; Bräutigam/Klindt (fn. 12), 1140; Kirn/Müller-Hengstenberg (fn. 2), 68; Horner/Kaulartz (fn. 92), 24.

117 Gerald Spindler, Zivilrechtliche Fragen beim Einsatz von Robotern, in: Hilgendorf (ed.), Robotik im Kontext von Recht und Moral (2014), 63–80, 72 ff., 78.

118 Horner/Kaulartz (fn. 92), 25.

119 On a convincing legal policy conception of corporate liability, Münchener Kommentar BGB/Wagner, 7th ed. 2017, BGB, § 831, Rn. 3 f.



#### 4. Gefährdungshaftung oder digitale Assistenzhaftung?

Aber *de lege ferenda* scheint die Gefährdungshaftung der Königsweg zu sein, auf dem man dem digitalen Autonomierisiko erfolgreich begegnen könnte.<sup>120</sup> Die meisten Autoren fordern auch mit großem Nachdruck eine entsprechende Gesetzgebung, die auf die digitalen Gefahren mit einem neuen Gefährdungshaftungstatbestand reagiert. Vorbilder dafür sind Gefährdungshaftungsnormen für Kraftfahrzeuge, für nukleare Anlagen, für gentechnische Verfahren und für Umweltschäden.

Doch muss hier ein fundamentales Missverständnis ausgeräumt werden. Die Leitprinzipien der Gefährdungshaftung können gar nicht als Vorbild dienen, da sie die Eigenheiten des digitalen Autonomierisikos schlicht verfehlen. Die These heißt: Nicht eine Haftung für den rechtmäßigen Einsatz gefährlicher Anlagen, sondern Haftung für rechtswidriges Fehlverhalten des autonom entscheidenden Softwareagenten ist als Grundprinzip einer strikten Haftung für digitales Handeln angemessen. Entsprechend wird hier der Vorschlag unterbreitet, statt Gefährdungshaftung eine „digitale Assistenzhaftung“ d. h. eine strikte Haftung des Prinzipals für Fehlverhalten des Softwareagenten einzuführen.

Zur Gefährdungshaftung besteht der folgende gravierende Unterschied. Gefährdungshaftung setzt ein beim Einsatz gefährlicher Sachen, der aber wegen seines gesellschaftlichen Nutzens erlaubt, also rechtmäßig ist.<sup>121</sup> Gefährdungshaftung ist also der seltene Fall einer Schadenshaftung trotz rechtmäßigen Verhaltens der Eigentümer/Nutzer/Betreiber. Gehaftet wird, wenn sich die typische Betriebsgefahr verwirklicht, also wenn Kausalabläufe fehlgelaufen sind. Bei Softwareagenten kommt es aber gerade nicht auf die Sachgefahr eines falsch funktionierenden Computers, also auf das Kausalrisiko, an, sondern auf das Entscheidungsrisiko, auf die andersgeartete Gefahr, dass sich dessen autonome Entscheidungen als Fehlentscheidungen herausstellen. Zurechnungsgrund ist nicht der Einsatz eines Objekts erhöhter Gefahr, sondern das rechtswidrige Verhalten des Algorithmus, den der Prinzipal rechtmäßig zu eigenem Nutzen eingesetzt hat.<sup>122</sup>

#### 4. Strict liability or digital assistance liability?

*De lege ferenda* seems to be the royal road to successfully countering the digital autonomy risk.<sup>120</sup> Most of the authors also insist strongly on a corresponding legislation that responds to the digital dangers with a strict liability. Examples of this are the hazardous liability standards for motor vehicles, nuclear installations, genetic engineering and environmental damage.

However, a fundamental misunderstanding must be dispelled. The guiding principles of strict liability cannot serve as a role model at all, since they simply fail to take account of the peculiarities of the digital autonomy risk. The thesis goes: Not liability for the lawful use of dangerous equipment, but rather liability for illegal misconduct of the autonomous, decision-making software agents is appropriate as a basic principle of vicarious liability for digital action. Accordingly, the proposal is presented here to introduce a “digital assistance liability” i.e. liability of the principal for misconduct by the software agent instead of strict liability.

The following fundamental difference exists with regard to strict liability. Dangerous things are subject to strict liability in their use, which is nevertheless permitted, i.e. lawful due to social benefits.<sup>121</sup> In other words, strict liability is the rare case of liability for damages despite the lawful conduct of the owner/user/operator. Liability occurs when the typical operational hazard is realized, i.e. when causal processes have gone awry. In the case of software agents, however, it is precisely not the hazard of a wrongly functioning computer, i.e. the causal risk, but rather the decision risk, the very different kind of risk that autonomous decisions will turn out to be wrong decisions. The reason for this is not the use of an object of increased risk, but the illegal behavior of the algorithm, which the principal has legitimately used for his own benefit.<sup>122</sup>

120 Europäisches Parlament (Fn. 3), 6; Günther (Fn. 17), 237 ff.; Schirmer (Fn. 23), 665; Gruber (Fn. 27), 198; Hanisch (Fn. 104, 2014), 35 f.; Spindler (Fn. 17), 775.

121 Locus classicus *Josef Esser*, Grundlagen und Entwicklung der Gefährdungshaftung: Beiträge zur Reform des Haftungsrechts und zu seiner Wiedereinordnung in die Gedanken des allgemeinen Privatrechts (1941). Zu einer soziologischen Interpretation *Luhmann* (Fn. 74), 69 f.

122 Zu dem entscheidenden Unterschied von Gefährdungshaftung und strikter Rechtswidrigkeitshaftung *Münchener Kommentar BGB/Wagner*, 7. Aufl. 2017, BGB, § 831, Rn. 4.

120 European Parliament (fn. 3), 6; Günther (fn. 17), 237 ff.; Schirmer (fn. 23), 665; Gruber (fn. 27), 198; Hanisch (fn. 104, 2014), 35 f.; Spindler (fn. 17), 775.

121 Locus classicus, *Josef Esser*, Grundlagen und Entwicklung der Gefährdungshaftung: Beiträge zur Reform des Haftungsrechts und zu seiner Wiedereinordnung in die Gedanken des allgemeinen Privatrechts (1941). On a sociological interpretation, *Luhmann* (fn. 74), Ch. 4, II.

122 On the decisive difference between absolute liability and strict unlawfulness liability, *Münchener Kommentar BGB/Wagner*, 7th ed. 2017 Edition, BGB, § 831, Rn. 4.

Weil die Autonomiegefahr digitaler Entscheidungen schlechterdings nicht mit der Betriebsgefahr der Gefährdungshaftung gleichgesetzt werden kann, müssen hier auch unterschiedliche Verantwortungsprinzipien und unterschiedliche Normen eingreifen. Haftungsnormen für Fehlentscheidungen autonomer Agenten können nicht auf das Kausalrisiko von Sachen gestützt werden, sondern sind auf das Entscheidungsrisiko von Akteuren zuzuschneiden. Auch an dieser Stelle realisiert sich die schon oben angesprochene Funktion der Personifizierung nicht-menschlicher Entitäten, dass sie die Kausalzurechnung durch eine Handlungszurechnung ersetzt.<sup>123</sup> Wenn das Privatrecht, wie hier vorgeschlagen, Softwareagenten als Stellvertreter und als Erfüllungsgehilfen, also als rechtlich handlungsfähig behandelt, dann ist es schlechthin ausgeschlossen, bei außervertraglichen Schädigungen mit einer bloßen Kausalhaftung zu operieren.

Kurz: *De lege ferenda* sollte in der Tat eine „digitale Assistenzhaftung“, d. h. eine strikte Haftung für Schädigungen durch autonome Softwareagenten, eingeführt werden, aber keinesfalls als Gefährdungshaftung für den rechtmäßigen Einsatz einer gefährlichen Anlage, sondern als Haftung eines Prinzipals für rechtswidrige Fehlentscheidungen seines Softwareagenten.<sup>124</sup> Praktisch relevant wird dieser grundsätzliche Unterschied für zwei Haftungsvoraussetzungen: (1) Anders als bei der Gefährdungshaftung wird für rechtswidriges Verhalten und zwar des Softwareagenten gehaftet. (2) Die Haftung erstreckt sich auf mehr als bloß auf Personen- und Sachschäden. Bisherige Regeln der Gefährdungshaftung anzuwenden, ginge also einerseits zu weit, weil sie auch bei rechtmäßigem Verhalten, das letztlich die Schädigungen verursacht, Haftung auslöst. Andererseits ginge sie nicht weit genug, weil sie regelmäßig nur für Personen- und Sachschäden Ersatz gewährt.

Bei der Gefährdungshaftung wird bekanntlich nicht nach Rechtswidrigkeit gefragt, bei Softwareagenten dagegen wird sie zum Dreh- und Angelpunkt der Haftung.<sup>125</sup> Es kommt entscheidend darauf an, ob die Handlung des Softwareagenten rechtswidrig war. Während es bei der Gefährdungshaftung ausreicht, wenn bloße Kausalität

Since the autonomy risk of digital decisions cannot, however, be equated with the causality hazard of strict liability, different principles of responsibility and different standards must also be developed here. Norms of liability for unlawful decisions by autonomous agents cannot be based on the causal risk of things, but must be tailored to the decision risk of actors. At this point, too, the function of personifying non-human entities realizes that it replaces causal attribution with an action attribution.<sup>123</sup> If private law, as proposed here, treats software agents as vicarious agents, i.e. as legally capable of acting, then it is absolutely impossible to operate with a mere causal liability in the case of non-contractual damages.

In short: *De lege ferenda* a “digital assistance liability”, i.e. strict liability for damages caused by autonomous software agents should indeed be introduced, but not as a liability for the legitimate use of a dangerous installation, but as a liability of a principal for unlawful decisions of his software agent.<sup>124</sup> This fundamental difference becomes practically relevant for two conditions of liability: (1) In contrast to strict liability, liability is assumed for unlawful behavior by the software agent. (2) The liability extends to more than just personal injury and property damage. In other words, it would go too far on the one hand, because it would also trigger liability in the event of conduct that simply causes the damage. On the other hand, it would not go far enough because it only provides compensation for personal injury and property damage.

Unlawfulness is not a condition for strict liability, but in the case of software agents it becomes the linchpin of liability.<sup>125</sup> It all depends on whether the software agent’s action was unlawful or not. While in the case of strict liability, it is sufficient if there is merely causality between operational hazard and damage, when using software agents it must be

<sup>123</sup> Dazu Teubner (Fn. 29).

<sup>124</sup> Ähnlich Hanisch (Fn. 104), 46 ff., der statt einer Gefährdungshaftung eine Haftung für „maschinelles Fehlverhalten“ fordert. Ähnlich auch mit dem rechtspolitischen Vorschlag einer Hilfspersonenhaftung für Roboter: Melinda Müller, Roboter und Recht: Eine Einführung, Aktuelle juristische Praxis (2014), 595–608, 601.

<sup>125</sup> Auch der Entwurf des Europäischen Parlaments verkennt die Besonderheit einer Softwarehaftung im Unterschied zur Gefährdungshaftung, da diese nur Kausalität zwischen schädlichem Verhalten des Computers und dem Schaden verlangt, Europäisches Parlament (Fn. 3), Rn. 27. Richtig dagegen Hanisch (Fn. 104, 2014), 46, der maschinelles Fehlverhalten zur Haftungsvoraussetzung macht.

<sup>123</sup> Teubner (fn. 29).

<sup>124</sup> Similarly Hanisch (fn. 104), 46 ff., who demands liability for “machine misconduct”, instead of strict liability. Similarly also with the proposal *de lege ferenda* of vicarious liability for robots, Melinda Müller, Roboter und Recht: Eine Einführung, Aktuelle juristische Praxis (2014), 595–608, 601.

<sup>125</sup> The European Parliament’s draft also fails to recognize the specificity of software liability, in contrast to strict liability, which only calls for a causal link between the harmful behavior of the computer and the damage incurred European Parliament (fn. 3), Rn. 27. Correctly, Hanisch (fn. 104, 2014), 46, who makes machine misconduct a prerequisite for liability.

zwischen Betriebsgefahr und Schaden besteht, muss beim Einsatz von Softwareagenten geprüft werden, ob dessen Verhalten rechtswidrig war oder nicht.

Der zweite wichtige Unterschied zur Gefährdungshaftung betrifft den Umfang der Haftung. Während Gefährdungshaftungsnormen regelmäßig die Haftung für Vermögensschäden ausschließen oder die Haftung auf nur wenige Rechtsgutsverletzungen beschränken, was angesichts des jeweils typischen Risikos auch gerechtfertigt ist, sind bei Softwareagenten solche Haftungsbeschränkungen nicht akzeptabel. Denn ihre typischen Risiken verwirklichen sich nicht bloß in den Kontexten, in denen gefährliche Anlagen Unfälle verursachen, sondern in all den Kontexten, in denen auch Menschen rechtswidrige Entscheidungen fällen. Wie bei schädigenden Handlungen natürlicher Personen müssen auch hier Vermögensschäden ausgeglichen werden, sofern das Handeln der Agenten ein Rechtsgut verletzt, gegen ein Schutzgesetz verstoßen hat oder sittenwidrig war. Und das macht einen großen Unterschied.

Auf welche Vermögensträger eine solche strikte Haftung für Computer-Fehlentscheidungen dann zuzurechnen ist, dafür sollte nach dem Vorbild der „Zurechnungs- und Haftungseinheit“, in der die Rechtsprechung bei Kfz-Gefährdungshaftung Fahrer, Halter, Versicherung zusammengefasst hat, verfahren werden.<sup>126</sup> Betreiber, Hersteller und Programmierer des Softwareagenten sollten zu einem ähnlichen Haftungsverbund zusammengeschlossen werden. Auch der Vorschlag von Hanisch, der Vorschlag einer abgestuften Haftung zwischen Betreiber und Hersteller, wonach primär der Betreiber haftet und sekundär der Hersteller, ist in diesem Zusammenhang überlegenswert.<sup>127</sup> Eine die Hersteller treffende Pflichtversicherung für die digitalen Risiken würde Härten für den Betreiber, besonders wenn sie nicht gewerbsmäßig handeln, abfedern.<sup>128</sup> Die Einführung von Höchstsummen entspräche der Versicherungslogik.

De lege ferenda sinnvoll wäre durchaus eine Generalklausel, die eine Veranlasserhaftung für autonome Algorithmen vorschreibt. Die Generalklausel müsste nur anordnen, dass für schadensverursachendes Handeln des Softwareagenten der Veranlasser einstehen muß, sofern das Handeln des Softwareagenten rechtswidrig war.<sup>129</sup> Versicherungslösungen, speziell Zwangsversicherungen stellen ihrerseits eine Änderung des Rechtsstatus der Softwareagenten dar. Sie ordnen eine Haftungsmasse ausschließlich den Handlungen der

checked whether their behavior was unlawful or not.

The second important difference to strict liability concerns the extent of liability. While strict liability standards regularly exclude liability for pecuniary loss which is justified in view of the typical risk, such liability limitations are not acceptable for software agents. Because their typical risks are not only realized in the contexts in which dangerous installations cause accidents, but also in all the contexts in which people make unlawful decisions. As in the case of damaging acts by natural persons, financial losses must be compensated for here as well, if the actions of the agents violated a legal right or violated a statutory rule or were against public policy. And that makes a big difference.

The question of which asset holders such strict liability for mistaken computer decisions can then be attributed should be based on the model of the “attribution and liability unit”, in which the case-law in the case of liability for hazardous vehicles has summarized drivers, owners and insurance companies.<sup>126</sup> Operators, manufacturers and programmers of the software agent should be bundled to a similar liability network. The proposal by Hanisch of a tiered liability between the operator and the manufacturer, according to which the operator is primarily liable and the manufacturer secondarily so, is also worth considering in this context.<sup>127</sup> A compulsory insurance policy for digital risks that affects manufacturers would cushion the burden of hardship for the operator, especially if they are not acting commercially.<sup>128</sup> The introduction of maximum sums would correspond to the insurance logic.

*De lege ferenda*, a general clause stipulating initiators’ liability for autonomous algorithms would make sense. The general clause would only have to provide that the initiators are responsible for damage-causing actions of the software agent, if the actions of the software agent were unlawful.<sup>129</sup> Insurance solutions, especially compulsory insurance, in turn represent a change in the legal status of software agents. They allocate a liability capital exclusively to the actions of the software agents, who thus do not become legal entities with their

126 Zech (Fn. 3), 204.

127 Hanisch (Fn. 104, 2014), 55 ff.

128 Europäisches Parlament (Fn. 3), Rn. 29.; Horner/Kaulartz (Fn. 92), 26.

129 Hanisch (Fn. 104, 2014), 46 ff., 54.

126 Zech (fn. 3), 204.

127 Hanisch (fn. 104, 2014), 55 ff.

128 European Parliament (fn. 3), Rn. 29.; Horner/Kaulartz (fn. 92), 26.

129 Hanisch (fn. 104, 2014), 46 ff., 54.

Softwareagenten zu, der damit zwar nicht zu einer juristischen Person mit Eigenvermögen wird, aber den Rechtsstatus eines „Haftungssubjekts mit „Versicherungsvermögen“ erhält. Nur eben dass dieses Vermögen nicht eigentumsmäßig zugeordnet wird, sondern sozialisiert wird, entweder auf die Versichertengemeinschaft oder gar auf die Gesamtgesellschaft.<sup>130</sup> In diesem Sinne verleihen Versicherungslösungen funktionale Äquivalente einer vollen Rechtssubjektivität. Praktisch wichtiger ist: Auch Versicherungslösungen implizieren die Handlungsfähigkeit der Softwareagenten. Wenn Versicherungen für die sogenannte Gefährdungshaftung eintreten, dann wäre es für sie ein nicht kalkulierbares Risiko, wenn sie für jeden kausal verursachten Schaden aufkommen müssten, ohne daß ein digitales Fehlverhalten vorliegt. Sie werden nur im Falle rechtswidrigen Handelns des Agenten Versicherungsschutz zu gewähren bereit sein, besonders dann, wenn die Ersatzpflicht über Personen- und Sachschäden hinausgeht.

Die vielfach geforderte Gefährdungshaftung für Softwareagenten ist also in Wahrheit nichts anderes als eine Variante einer „Gehilfenhaftung“, die notwendig rechtswidriges Verhalten des Agenten voraussetzt und deren Umfang nicht auf Personen- und Sachschäden begrenzt ist. Sie würde – implizit oder explizit – eine beschränkte Rechtsfähigkeit der Softwareagenten anerkennen.

#### IV.

#### HYBRIDE: VERBUNDRISIKO

##### 1. Mensch-Maschinen Assoziationen

Als Antwort auf das Autonomierisiko wurden bis hierher dogmatische Lösungen vorgeschlagen, in denen Algorithmen mit beschränkter Rechtsfähigkeit ausgestattet und als eigenständig entscheidende Stellvertreter oder Hilfspersonen behandelt werden. Damit bewegt man sich noch auf einigermaßen sicherem Boden der Rechtsdogmatik. Sehr viel unsicheres Gelände aber betritt man, wenn man den Verbund von Menschen und Algorithmen selbst als Regelungsgegenstand ins Visier nimmt. Der Anlass dafür ist, dass die bisher diskutierte Lösung, den Softwareagenten beschränkte Rechtsfähigkeit zuzuerkennen, ebenso wie die Lösung der traditionellen Dogmatik, das Softwareverhalten einfach als Menschenverhalten zu fingieren, auf eine problematische Alternative hinführt. Einerseits: Ausschließlich den beteiligten Menschen Rechte und Pflichten zuzuschreiben – wie es die traditionelle Dogmatik tut –, erfasst die digitalen

<sup>130</sup> Zur sozialversicherungsrechtlichen Lösung *Janal* (Fn. 3), 157.

own funds, but obtain the legal status of a “liability subject with insurance assets”. Only that these assets are not attributed as property, but are rather socialized, either to the community of insured persons or even to the entire society.<sup>130</sup> In this sense, insurance solutions lend functional equivalents to full legal subjectivity. More importantly, insurance solutions also imply the action capacity of software agents. If insurances are providing cover for strict liability, it would be an incalculable risk for them if they had to pay for every damage that came about causally, without the presence of digital misconduct. They will only be prepared to provide insurance cover in the event of unlawful conduct by the agent, especially if the obligation to pay compensation goes beyond personal injury and property damage.

The frequently called for strict liability for software agents is therefore in reality nothing more than a variant of an “assistant liability”, which necessarily presupposes unlawful behavior of the agent and whose extent is not limited to personal injury and property damage. It would implicitly or explicitly recognize a limited legal personhood of the software agents.

#### IV.

#### HYBRIDS: ASSOCIATION RISK

##### 1. Human-Machine Associations

In response to the digital autonomy risk, legal solutions have so far been proposed in which algorithms have limited legal personhood and are treated as autonomously decision-making representatives or auxiliaries. This is still a reasonably secure area for legal doctrine. However, one enters a perilous terrain when focusing on the human-algorithm association as a regulatory object. The reason for this is that both the solution of granting limited legal capacity to software agents, and the solution of simply dressing up software behavior as human behavior, leads to a problematic alternative. On the one hand, to attribute rights and obligations exclusively to the human actors involved does not adequately capture digital risks, because it ignores the autonomous role of the machines, spans almost with necessity the duties of the people involved and reduces the protection of the machines with regard to rights. On the other hand: The actions – as pro-

<sup>130</sup> On solutions under social security law, *Janal* (fn. 3), 157.

Risiken nicht ausreichend. Denn es ignoriert die autonome Rolle der Automaten, überspannt beinahe mit Notwendigkeit die Pflichten der beteiligten Menschen und reduziert in Bezug auf Rechte den Schutz der Automaten. Andererseits: Die Handlungen – wie hier vorgeschlagen – ausschließlich den Automaten zuzurechnen, ist auch nicht problemlos, denn man behandelt dann den Beitrag der Menschen im Mensch-Maschinen-Verbund nur in individualrechtlichen Kategorien von Stellvertretung und Vollmacht, was den Eigenheiten des Verbunds selbst nicht gerecht wird.

Auch hier bietet wieder Bruno Latours Netzwerk-Aktor-Theorie einen Ausweg. Sie realisiert die Mensch-Maschinen-Interaktion realitätsgerecht als Kollektivphänomen eigener Art: Nach Latour besitzen „Aktanten“ nicht nur eine Sprache und einen widerständigen Körper, sondern auch das Vermögen, „Assoziationen“ zu bilden.<sup>131</sup> Die Fähigkeit zum gesellschaftlichen Handeln verdanken sie nicht nur ihrem individuellen Status als „Aktanten“, sondern darüber hinaus ihrer Mitgliedschaft in „Hybriden“, d. h. in Assoziationen von menschlichen Akteuren und nicht-menschlichen Aktanten. Damit entsteht – ebenso wie in jeder Assoziation – ein „Ressourcenpool“, der im Falle der Hybriden die beschränkten Handlungskapazitäten der Aktanten mit den kommunikativen Fertigkeiten realer Menschen kombiniert. „Die psychosystemischen Kompetenzdefizite der Non-Humans werden durch die verteilte Intelligenz der Sozialsysteme zuverlässig kompensiert.“<sup>132</sup> Und es ist gerade die Verbindung von menschlichen und nicht-menschlichen Handlungskapazitäten innerhalb der Hybriden, die es nicht-menschlichen Entitäten erlaubt, auch an komplexeren politischen Verhandlungen, wirtschaftlichen Transaktionen und juristischen Beziehungen teilzunehmen.<sup>133</sup>

Ähnlich wie eine formale Organisation entwickelt der Mensch-Maschinen-Verbund selbst eine phänomenale Innenperspektive, eine Selbstwahrnehmung als ein „lebendiger“ Prozess, eine eigene Hierarchie von Präferenzen, eigene soziale Bedürfnisse und politische Interessen, die weder auf die Eigenschaften der beteiligten Menschen noch auf die der Algorithmen reduziert werden können. Die Interaktionen innerhalb des Verbundes machen die Hybride mit körperschaftlichen Akteuren vergleichbar. Die beteiligten Akteure handeln nicht für sich selbst, sondern „für“ den Hybriden als eine

posed here – to be attributed exclusively to the machines are also not without problems, because then the contribution of the people in the human-machine association is only dealt with in individualistic legal categories of agency law, which does not do justice to the peculiarities of the human-computer association itself.

Here again Bruno Latour’s actor-network theory offers a way out. It conceives of the human-machine interaction as a collective phenomenon of its own kind: According to Latour, “actants” possess not only a language and a resistant body, but also the ability to form “associations”.<sup>131</sup> They owe their participation in society not only to their individual status as “actants”, but also to their membership in “hybrids”, i.e. to associations of human actors and non-human actants. As in any association, this creates a resource pool, which, in the case of hybrids, combines the limited capacity of the actants to act with the communicative skills of real people. “The psychosystemic deficits of the non-human’s competence are reliably compensated for by the distributed intelligence of the social systems.”<sup>132</sup> And it is precisely the combination of human and non-human capacities for action within the hybrids that allows non-human entities to participate in more complex political negotiations, economic transactions and legal activities.<sup>133</sup>

Similar to a formal organization, the human-machine association itself develops a phenomenal inner perspective, a self-perception as a “living” process, its own hierarchy of preferences, its own social needs and political interests that cannot be reduced to the characteristics of the human actors involved or to those of the algorithms. The interactions within the network make the hybrids comparable with corporate actors. The actors involved do not act on their own behalf, but “for” the hybrids as an emergent entity, as an association of people and non-humans. They do it in the same way as manag-

131 Latour (Fn. 36), 104.

132 So Lorentzen, Luhmann Goes Latour: Zur Soziologie hybrider Beziehungen, in: Rammert/Schulz-Schaeffer (Hg.), Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik (2002), 101–118, 110.

133 Eingehend zur Kooperation von Menschen und Algorithmen Misselhorn (Fn. 49); Anita Fiebich/Nhung Nguyen/Sarah Schwarzkopf, Cooperation with Robots? A Two-Dimensional Approach, in: Misselhorn (Hg.), Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation (2005), 25–44.

131 Latour (fn. 36), 70 ff.

132 So Lorentzen, Luhmann Goes Latour: Zur Soziologie hybrider Beziehungen, in: Rammert/Schulz-Schaeffer (ed.), Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik (2002), 101–118, 110.

133 On the cooperation of humans and algorithms, Misselhorn (fn. 49); Anita Fiebich/Nhung Nguyen/Sarah Schwarzkopf, Cooperation with Robots? A Two-Dimensional Approach, in: Misselhorn (ed.), Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation (2005), 25–44.

emergente Einheit, als Assoziation von Menschen und Nicht-Menschen. Sie tun es in der gleichen Weise wie auch Manager im Unternehmen, die nicht im eigenen Namen handeln, sondern als „Agenten“ in Vertretung für ihren „Prinzipal“, für das Unternehmen selbst als soziales System. Und dort gibt es Interessen- und Orientierungskonflikte zwischen den Mitgliedern und zwischen ihnen und dem Verbund, wie sie angesichts des bekannten *Agency*-Problems bei korporativen Akteuren diskutiert werden, für deren Lösung zahlreiche Rechtsnormen entwickelt worden sind. Solche Konflikte existieren auf ähnliche Weise auch in den Assoziationen von Menschen und Nicht-Menschen. Und vergleichbare institutionelle Regulierungen – z. B. die Formalisierung der Pflichten und Verantwortlichkeiten des Geschäftsführers, die *Ultra-Vires*-Lehre, die Prüfung der Repräsentativität bei der Sammelklage – dienen dazu, das *Agency*-Problem einzudämmen, welches auch in den Mensch-Maschinen-Assoziationen auftaucht. In der Tat werden in der Rechtsdogmatik zum Electronic Contracting ähnliche Lösungsmöglichkeiten bezüglich der „Prinzipal-Agenten-Beziehung“ zwischen dem vertragsschließenden Computer und den menschlichen Vertragsparteien diskutiert.<sup>134</sup>

Dies ist eine kollektivrechtliche Sicht, die das Recht aus der nicht unproblematischen individualrechtlichen Alternative befreit, die fragliche Handlung entweder ausschließlich dem Menschen oder ausschließlich dem Algorithmus zurechnen zu müssen. Im Gegensatz zur Individualzurechnung ist sie in der Lage, das emergente Phänomen des Mensch-Maschinen-Verbundes auch rechtlich einzufangen.<sup>135</sup> Denn sie berücksichtigt zum einen die Emergenz in den Innenbeziehungen, die jenseits der Existenz der Binnenakteure die überpersonale Einheit ihres Verbundes erzeugt. Zum anderen wird sie der Emergenz in den Außenbeziehungen gerecht: der Mensch-Algorithmus-Verbund als eigenständiges Handlungssystem kommuniziert selbst mit Dritten, und es kommunizieren nicht mehr nur die beteiligten Menschen oder Algorithmen. Den hier auftretenden Risiken, die aufgrund der fast nicht mehr auflösbaren Verflechtung der Einzelhandlungen von Menschen und Algorithmen entstehen, lässt sich besser begegnen, wenn man die Mensch-Algorithmen-Assoziation, den Hybrid als solchen, als gemeinsamen Zurechnungspunkt für Handlungen, Rechte und Pflichten identifiziert.

ers in a company who do not act in their own name, but as “agents” acting on behalf of their “principals”, i.e. for the company itself as a social system. And there are conflicts of interest between the members and between them and the association, as discussed in the light of the well-known agency problem faced by corporate actors, for which the solution has been developed in numerous legal norms. Such conflicts also exist similarly in the associations of humans and non-humans. And comparable institutional regulations – e.g. the formalization of the duties and responsibilities of the managing director, the *Ultra-Vires* doctrine, the examination of representativeness in class action lawsuits – serve to contain the agency problem, which also arises in human-machine associations. In fact, in the legal debate on electronic contracting, similar solutions are being discussed with regard to the “principal-agent relationship” between the contracting computer and the human contracting parties.<sup>134</sup>

This is a collectivist view that frees the law from the not unproblematic individualist alternative of assigning the act in question either exclusively to humans or exclusively to the algorithm. In contrast to individual attribution, it is also capable of doing justice to the emergent phenomenon of the human-machine association.<sup>135</sup> On the one hand, it takes into account the emergence in internal relations which, beyond the existence of several individual actors, creates the unity of their association. On the other hand, it does justice to the emergence in external relations: the human-algorithm association as an independent system communicates itself with third parties and it is no longer only the people or algorithms involved to whom communication is attributed. The risks arising from the almost inextricable interweaving of the individual actions of humans and algorithms can be better counteracted by identifying the human-algorithms association, the hybrid as such, as a common point of attribution for actions, rights and obligations.

134 Juanda L. Daniel, *Electronic Contracting under the 2003 Revisions to Article 2 of the Uniform Commercial Code: Clarification or Chaos?*, Santa Clara Computer & High Technology Law Journal 20 (2004), 319–346, 344 ff.

135 Kersten (Fn. 17), 4 ff.; Ingold (Fn. 17), 220 ff.

134 Juanda L. Daniel, *Electronic Contracting under the 2003 Revisions to Article 2 of the Uniform Commercial Code: Clarification or Chaos?*, Santa Clara Computer & High Technology Law Journal 20 (2004), 319–346, 344 ff.

135 Kersten (fn. 17), 4 ff.; Ingold (fn. 17), 220 ff.

## 2. Hybride als Rechtssubjekte?

Sollte man dann auch rechtsdogmatisch echte Hybrid-Lösungen erwägen? Sollte man die Rechtskonstruktion eines Mensch-Maschinen-Verbundes einführen? Dies ist eine durchaus ernst zu nehmende, wenn auch bisher noch nicht ausprobierte Alternative. Die einzelnen rechtsgeschäftlichen Handlungen der Softwareagenten und die der beteiligten Menschen würden, ohne beide im Einzelnen trennen zu müssen, zu einer einheitlichen Handlung des Mensch-Maschinen-Verbundes zusammengezogen und würden sowohl Rechtsbindungen als auch Haftungsansprüche erzeugen. Im Gegensatz zum individualistischen Recht der Stellvertretung, das die Einzelhandlungen von Prinzipal und Agent klar voneinander trennt und den Prinzipal zum Vertragspartner erklärt, würde hier der Mensch-Maschinen-Verbund selbst zum eigentlichen Vertragspartner. Für vertragliche und außervertragliche Haftung würden die einzelnen Haftungsvoraussetzungen zusammengesetzt aus ihrer beider Verhalten, ohne dass ihre Einzelbeiträge mühsam und häufig willkürlich auseinandergerechnet werden müssten, wie das in der Haftung für Hilfspersonen notwendig wäre. Eine Wissenszurechnung schließlich fände weder ausschließlich auf den Computer noch ausschließlich auf den Menschen statt, sondern stets auf den Verbund. Auf Vermögen würde haftungsrechtlich zugegriffen, wo immer es sich im Hybriden befindet.

Jedoch: Die Hybride tatsächlich als rechtliche Einheit zu konstruieren, würde bedeuten, dass das Recht eine völlig neuartige gesellschaftsrechtliche Konstruktion, die Mensch-Maschinen-Assoziation, schaffen müsste.<sup>136</sup> Dies wäre in der Tat ein wagemutiger – um nicht zu sagen: waghalsiger – Schritt, auf den die heutige Privatrechtsdogmatik kaum vorbereitet ist. Algorithmen als vollwertige Mitglieder einer neuartigen Assoziation? Dieser Schritt wäre noch in einem viel stärkeren Maße eine Rechtsfortbildung extra legem als es der ohnehin schon recht kühne Schritt ist, den Softwareagenten als Repräsentanten ihres menschlichen Prinzipals Teilrechtsfähigkeit zuzuerkennen. Obwohl die kollektivrechtliche Verbundlösung der individualrechtlichen Repräsentantenlösung der Realität der Mensch-Maschinen-Interaktion letztlich eher angemessen ist, dürften Rechtsprechung und Lehre es wohl vorziehen, statt dem Verbund sämtliches Verhalten einheitlich zuzurechnen, das Verhalten der einzelnen Partizipanten, also der

<sup>136</sup> So die tastenden Überlegungen von Gruber (Fn. 27), 158; Allen/Widdison (Fn. 17), 40 mit Verweis auf die soziale Realität der Hybride. Für eine moralische Verantwortung von Mensch-Maschinen-Verbänden als „responsibility networks“ spricht sich Christian Neuhäuser, Some Sceptical Remarks Regarding Robot Responsibility and a Way Forward, in: Misselhorn (Hg.), *Collective Agency and Cooperation in Natural and Artificial Systems* (2015), 131–147, 143 f., aus.

## 2. Hybrids as legal entities?

Should legal doctrine introduce the legal construction of a human-machine association? This is a serious option, even if it has not yet been tried out. The individual legal acts of the software agents and those of the people involved, without having to separate them individually, would be merged into a unified act of the human-machine association and would create both legal ties and liability claims. In contrast to the individualistic law of agency, which clearly separates the individual actions of principals and agents and declares the principal to be the contractual partner, the human-machine association itself would become the actual contractual partner. For contractual and non-contractual liability, the individual preconditions for liability would be composed of their conduct, without their individual contributions having to be painstakingly and often arbitrarily calculated apart, as it would be necessary in the case of liability for auxiliary persons. Finally, knowledge allocation would not exclusively be based on the computer or on the human actor, but always on the association. Assets would be accessed under liability law wherever they are located in the hybrid.

However, to actually construct the hybrid as a legal entity would mean that the law would have to create a completely new kind of corporate entity, the human-machine association.<sup>136</sup> This would indeed be a bold – if not daring – step, for which today’s private law is hardly prepared. Algorithms as full members of a novel association? To a much greater extent, this construction would be more radical than granting software agents partial legal personhood as representatives of their human principal. Although the collectivist legal solution is ultimately more appropriate to the reality of human-machine interaction, courts and doctrine are likely to prefer, instead of attributing all behavior uniformly to the association, to choose the individualist agency law solution.

<sup>136</sup> Gruber (fn. 27), 158; Allen/Widdison (fn. 17), 40 referring to the social reality of the hybrid. Christian Neuhäuser, Some Sceptical Remarks Regarding Robot Responsibility and a Way Forward, in: Misselhorn (ed.), *Collective Agency and Cooperation in Natural and Artificial Systems* (2015), 131–147, 143 f. advocates a collective moral responsibility of man-machine associations as “responsibility networks”.

Einzelmenschen und der Softwareagenten, rechtlich als eine Art Repräsentation individuell Handelnder zu konstruieren.

Wie dem auch sei, selbst wenn es bei der individualrechtlichen Lösung verbleibt, ist auch dann die Einheit des Mensch-Algorithmus-Verbundes juristisch nicht irrelevant; sie wird nun als rechtlich bedeutsame Hintergrundsannahme wirksam. Ein interessantes Vorbild, mit einer solchen *unitas duplex* umzugehen, bietet schon das geltende Vertragsrecht mit dem Vertragszweck, dem gemeinsamen Zweck der Individualparteien in der Austauschbeziehung des Schuldvertrages. Bekanntlich wird über den Vertragszweck der Vertrag nicht selbst als Kollektiveinheit konstruiert, sondern weiterhin streng individualistisch als Beziehung zwischen den einzelnen Vertragspartnern verstanden. Aber der Vertragszweck, verstanden als der gemeinsame Austauschzweck der Parteien, bringt die übergreifende Einheit der Vertragsbeziehung anders zur Geltung. Er erzeugt konkrete rechtliche Wirkungen – für die Auslegung der Parteierklärungen, für die Normierung von Treupflichten und für die Entscheidung in Streitfällen, ob und in welcher Intensität ein Vertragsbruch vorliegt. Ja, in Dauerschuldverhältnissen, relationalen Verträgen und ganz besonders in Vertragsverbänden, z. B. in Zulieferer- und Vertriebsnetzen, schiebt sich dieser Vertragszweck immer stärker in den Vordergrund – nun unter den Titeln Verbundzweck, Finalnexus oder Netzzweck –, und dies stets, ohne dass gesellschaftsrechtliche Konstruktionen gewählt werden müssen.<sup>137</sup> In ähnlicher Weise müssten die Mensch-Computer-Verbände nicht selbst mit rechtssubjektiven Qualitäten ausgestattet werden, aber ein „Verbundzweck“ könnte ihrem hybriden Charakter gerecht werden, eine rein individualistische Konstruktion korrigieren und die Einheit des Mensch-Maschinen-Verbundes auch rechtlich zur Geltung bringen.

Letztlich sind die individualrechtlichen und die kollektivrechtlichen Lösungen gar nicht als einander ausschließende Alternativen zu verstehen. Sie könnten durchaus sinnvoll nebeneinander bestehen. Die Agenten-Prinzipal-Lösung wäre dann angebracht, wenn die Algorithmen im Sozialleben als klargeschnittene Akteure auftreten. Wenn sie dagegen in einen dichten Interaktionszusammenhang mit menschlichen Akteuren eingebettet sind, dürfte die Verbundlösung vorzuziehen sein. Beide Lösungen hätten jedoch gemeinsam zur notwendigen Voraussetzung, dass sie beschränkte Rechtsfähigkeit an unterschiedliche nicht-menschliche Akteure – an Aktanten oder an Hybride – verleihen müssten.

<sup>137</sup> Zur neuartigen Kategorie des Netzzwecks: Gunther Teubner, Netzwerk als Vertragsverbund: Virtuelle Unternehmen, Franchising, Just in Time in sozialwissenschaftlicher und juristischer Sicht (2004), 148 ff.

Be that as it may, even if the agency law is preferred, the unity of the human-algorithm association will not be legally irrelevant; it now becomes effective as a legally significant background assumption. An interesting model for dealing with such a *unitas duplex* is already offered by the current contract law with the purpose of the contract, the common purpose of the individual parties in the exchange relationship. It is well known that the purpose of the contract is not to construct the contract itself as a collective unit, but to continue to be understood strictly individualistically as a relationship between the individual contracting parties. But the purpose of the contract, understood as the common purpose of exchange between the parties, brings the overarching unity of the contractual relationship to bear differently. It produces concrete legal consequences – for the interpretation of contractual declarations, for fiduciary duties and for the decision in disputes as to whether and to what extent a breach of contract has occurred. Moreover, in long-term contracts, relational contracts and especially in contractual networks, e.g. in supplier and distribution networks, this purpose of the contract is becoming more and more important – now under the titles “association purpose”, “final nexus” or “network purpose” – without having to choose corporate law constructions.<sup>137</sup> Similarly, the human-computer associations would not have to be endowed with legal capacities themselves; rather an “association purpose” could do justice to their hybrid character, correct a purely individualistic construction and bring the unity of the human-machine association to bear in legal terms.

Ultimately, both solutions are not to be understood as mutually exclusive alternatives. They could very well exist side by side. The agent-principal solution would be appropriate if the algorithms were to act in social life as clearly defined actors. If, on the other hand, they are embedded in a dense interaction context with human actors, the associational solution should be preferred. However, the two solutions have in common the necessary precondition that they confer limited legal personhood to non-human actors – actants or hybrids.

<sup>137</sup> On the new category of the network purpose, Gunther Teubner, Networks as Connected Contracts (2011), 184 ff.



V.  
MULTI-AGENTEN-SYSTEME:  
VERNETZUNGSRISIKO

Die Rechtstechnik der Personifizierung von Aktanten oder Hybriden gerät jedoch dann an ihre Grenzen, wenn autonome Algorithmen in einem Multi-Agenten-System vernetzt sind. Personifizierung hat ein bestimmbares Substrat noch bei klar identifizierbaren Softwareagenten oder eindeutig abgrenzbaren Mensch-Maschinen-Assoziationen, nicht aber mehr bei komplexen Computervernetzungen.<sup>138</sup> Das hier auftretende Vernetzungsrisiko zerstört Annahmen über Individualität von Akteuren, die für die Zurechnung von Handlung und Verantwortung konstitutiv sind.<sup>139</sup> Sowohl der Handlungsträger als auch die Kausalzusammenhänge sind dann schwierig, wenn nicht unmöglich, aufzuklären.<sup>140</sup>

Spiecker zeigt in aller Deutlichkeit die Verantwortungslücke des geltenden Rechts. Ihr Vorschlag einer Pro-Rata-Haftung aller an der Vernetzung Beteiligten, die für das rechtswidrige Verhalten der vernetzten Algorithmen einstehen müssten,<sup>141</sup> ist aber kaum praktikabel. Angesichts der beinahe unendlichen Vernetzung von Algorithmen im digitalen Raum ist eine Abgrenzung der Haftungssubjekte praktisch unmöglich. Hinzukommen kaum vertretbare Ermittlungskosten: „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.“<sup>142</sup>

Ein Ausweg wäre, rechtliche Verantwortung nicht mehr „Personen“ als Handlungsträgern, sondern nur noch identifizierbaren Handlungen, also einer „anonymen Matrix“ von sozialen und digitalen Prozessen selbst (und nicht auf Strukturen), zuzurechnen. Die eigentlichen Zurechnungspunkte für Verantwortung sind dann autonome Entscheidungen und nicht mehr die Entscheidungsträger. Letztlich wären nicht Menschen, nicht Organisationen, nicht Netzwerke, nicht Softwareagenten, nicht Algorithmenvernetzungen, sondern die Entscheidungen selbst in die Verantwortung zu nehmen.

In einer solchen Situation bliebe nichts anderes übrig, als dass das Recht von sich aus autonom abgegrenzte Risiko-Pools zur Eingrenzung dieser Handlungszusammenhänge konstruiert, und zwar

V.  
MULTI-AGENT SYSTEMS:  
NETWORK RISK

The legal technique of personifying actants or hybrids, however, reaches its limits when many autonomous algorithms are interconnected within a multi-agent system. Personification needs a determinable substrate for clearly identifiable software agents or definite human-machine associations, but it does not work in a situation of complex computer interconnections.<sup>138</sup> The network risk that is realized here destroys assumptions about the individuality of actors which are constitutive for the attribution of action and responsibility.<sup>139</sup> Both the actor and the causal relationships are difficult, if not impossible, to identify.<sup>140</sup>

Spiecker clearly shows the responsibility gap of the current law. However, her proposal for a pro-rata liability of all those involved in the network, who would have to stand up for the unlawful behavior of the networked algorithms,<sup>141</sup> is hardly practicable. In view of the almost infinite interconnection of algorithms in the digital space, it is practically impossible to delimit liability subjects. In addition, there are huge investigation costs: “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.”<sup>142</sup>

One way out of this dilemma would be to attribute legal responsibility to identifiable actions, i.e. to an “anonymous matrix” of social and digital processes themselves, rather than to “persons” as actors. The actual attribution points for responsibility are then autonomous decisions and no longer the decision-makers. Ultimately, it would not be people, organizations, networks, software agents, algorithms, but rather the decisions themselves that would have to be made responsible.

In such a situation, the law itself would need to construct risk pools for the delimitation of these interrelationships, especially if no factual indications of personification are identifiable in the digital

138 *Spiecker* (Fn. 7), 701.

139 *Susanne Beck*, Dealing with the Diffusion of Legal Responsibility: The Case of Robotics, in: Battaglia/Mukerji/Nida-Rümelin (Hg.), *Rethinking Responsibility in Science and Technology* (2014), 167–182; *Floridi/Sanders* (Fn. 6), 205 ff.

140 *Zech* (Fn. 3), 170.

141 *Spiecker* (Fn. 7), 703.

142 *Koops et al.* (Fn. 19), 553.

138 *Spiecker* (fn. 7), 701.

139 *Susanne Beck* (2014) “Dealing with the Diffusion of Legal Responsibility: The Case of Robotics”, in: Fiorella Battaglia, Nikil Mukerji and Julian Nida-Rümelin (ed.) *Rethinking Responsibility in Science and Technology* (2014), 167–182; *Floridi/Sanders* (fn. 6), 205 ff.

140 *Zech* (fn. 3), 170.

141 *Spiecker* (fn. 7), 703.

142 *Koops et al.* (fn. 19), 553.

gerade auch dann, wenn im digitalen Raum keine faktischen Anhaltspunkte für Personifizierung identifizierbar sind.<sup>143</sup> Hier verliese das Recht endgültig die Akteursperspektive, weil es nicht einmal mehr nach Einzel- oder Kollektivakteuren sucht, sondern sich auf die riskanten Entscheidungen als solche konzentriert. Es macht Handlungsketten selbst verantwortlich, ohne sich um deren Eigenschaft als organisierte Willensbildungseinheiten zu kümmern.

Der entscheidende Unterschied einer solchen Risikohaftung zu bekannten Formen der Organisationshaftung besteht darin, dass das Haftungsrecht sich nicht mehr auf in der Realität bestehende organisatorische Arrangements oder wenigstens auf Kooperationszusammenhänge bezieht, sondern selbst neuartige Risikonetzwerke definiert, um nicht zu sagen: dekretiert. Und sobald die Handlungen individueller, kollektiver Akteure oder die Kalkulationen digitaler Akteure in einen solchen Raum geraten, würden sie alle zu „Zwangsmitgliedern“ eines solchen Risikopools – also nicht kraft privatautonomer Entscheidung, sondern kraft autoritativer Anordnung des staatlichen Rechts. Freilich gibt es noch einen privatautonomen Anteil am Geschehen, er ist aber auf die Eintrittsentscheidung beschränkt. Sie sind dann in der Kollektivhaftung ohne Rücksicht auf ihren Willen und ohne Rücksicht auf den Kausalzusammenhang zwischen den von ihnen individuell verantwortbaren Handlungen und dem eingetretenen Schaden.<sup>144</sup>

Der Risikopool würde nicht mehr durch kooperative, organisatorische oder technische Strukturen bestimmt. Er sollte vielmehr als „digitaler Problembereich“ definiert werden, dessen Grenzen durch die Eignung für kollektives Risikomanagement zu bestimmen sind. Letztlich entscheidend sind weder Kausalzusammenhänge noch vorgegebene kooperative Strukturen – so wichtig beides im Einzelfall auch sein mag –, sondern zentrales Kriterium sollte die Fähigkeit des Pools zum Risikomanagement sein. Zugegeben, das wäre „opportunistische“ Zurechnung,<sup>145</sup> in diesem Fall aber nicht kausaler, sondern kollektiver Art. Für den Zweck der kollektiven

space.<sup>143</sup> Here, the law would finally leave the actor perspective, because it no longer looks for individual or collective actors but rather focuses on the risky decisions as such. It makes chains of actions responsible for their consequences without caring for organized decision centers.

The decisive difference between such risk liability and known forms of organizational liability lies in the fact that liability law no longer refers to organizational arrangements that exist in reality or at least refers to cooperative relationships, but rather defines, not to say decrees, new types of risk networks. And as soon as the actions of individual, collective actors or the calculations of digital actors move into such a space, they all become “compulsory members” of such a risk pool—by the authoritative order of state law. Admittedly, there is still a voluntary part of the action, but it is limited to the decision on entry. They then become collectively liable irrespective of their intention and without regard to the causal relationship between the actions they are individually responsible for and the damage incurred.<sup>144</sup>

The risk pool would no longer be determined by cooperative, organizational or technical structures. Rather, it should be defined as a “digital problem area”, the limits of which should be determined by the suitability for collective risk management. Ultimately, neither causal connections nor pre-defined cooperative structures – however important both are in individual cases – are decisive, but the central criterion should be the pool’s ability to manage risk. Admittedly, this would amount to “opportunistical” attribution.<sup>145</sup> For the purpose of collective liability, it is the law that identifies by its own authority concrete digital risk contexts of the offline and online

143 In diese Richtung, das „behaviour“, also die Handlungen und nicht die Handlungsträger, selbst als Zurechnungspunkte zu definieren, denkt anscheinend, wenn vielleicht auch etwas vorsichtiger, *Hildebrandt* (Fn. 37), 26: „Because the agents may be distributed on and possibly mobile between different hardware applications and because as a multi-agent system it is capable of changing shape (polymorphous), it is not always easy to identify where the emerging agent is located and what is and is not a part of it at any point in time. However, in so far as the emergent behaviours of the system allow its identification as a unity of action, it can be qualified as an agent, whatever the underlying embodiment“.

144 Zu parallelen Überlegungen im Umwelthaftungsrecht EPA, Superfund Enforcement: 35 Years of Protecting Communities and the Environment, <https://www.epa.gov/enforcement/superfund-enforcement-35-years-protecting-communities-and-environment>; *Gunther Teubner*, Die unsichtbare „Cupola“: Kausalitätskrise und kollektive Zurechnung. In: Lübbe (Hg.) Kausalität und Zurechnung (1994), 91-143; *Gerhard Wagner*, Kollektives Umwelthaftungsrecht auf genossenschaftlicher Grundlage (1990).

145 *Luhmann* (Fn. 74), 129.

143 *Hildebrandt* (fn. 37), 26 as well defines actions and not actors as points of attribution, although perhaps more cautiously: “Because the agents may be distributed on and possibly mobile between different hardware applications and because as a multi-agent system it is capable of changing shape (polymorphous), it is not always easy to identify where the emerging agent is located and what is and is not a part of it at any point in time. However, in so far as the emergent behaviours of the system allow its identification as a unity of action, it can be qualified as an agent, whatever the underlying embodiment.”

144 On concurrent considerations in environmental liability law, EPA, Superfund Enforcement: 35 Years of Protecting Communities and the Environment, <https://www.epa.gov/enforcement/superfund-enforcement-35-years-protecting-communities-and-environment>; *Gunther Teubner*, Die unsichtbare „Cupola“: Kausalitätskrise und kollektive Zurechnung. In: Lübbe (Hg.) Kausalität und Zurechnung (1994), 91-143; *Gerhard Wagner*, Kollektives Umwelthaftungsrecht auf genossenschaftlicher Grundlage (1990).

145 *Luhmann* (fn. 74), Ch. 6, V.

Haftung identifiziert kraft autoritativer Anordnung das Recht konkrete digitale Risikozusammenhänge der Offline- und der Online-Welt mit dem Hintergedanken, ein soziales Gebilde zu schaffen, das diese Risiken einigermaßen präventiv beherrschen oder wenigstens zur Schadensabwicklung herangezogen werden kann.

Dies Risikomanagement betrifft erstens die Abwicklung bereits entstandener Schäden. Das Recht gleicht in Fällen mehrfacher Kausalität die Schäden so aus, dass es einen ausreichenden finanziellen Pool schafft, der die Verluste abdeckt und das Risiko verteilt („*deep pocket*“, „*risk spreading*“). Zweitens – und das ist womöglich wichtiger – bedeutet Risikomanagement die kollektive Steuerung zukünftigen Verhaltens. Das Recht zieht die Grenzen des Risikopools derart, daß eine realistische Basis für eine aktive und gemeinsame Prävention von Risiken in Gebieten geschaffen wird, wo Schädigungsprobleme konzentriert sind. Unter beiden Gesichtspunkten isoliert das Recht den sozialen Bereich kollektiver Verantwortlichkeit so, daß sich eine funktionsfähige digitale Technologie zur Bewältigung der digitalen Vernetzungsrisiken entwickeln kann.

## VI. ERGEBNIS

Die drei geschilderten digitalen Risiken – Autonomierisiko, Verbundrisiko, Vernetzungsrisiko – stellen das Privatrecht vor die Herausforderung, den Rechtsstatus von autonomen digitalen Informationssystemen neu zu bestimmen. Freilich nicht in dem Sinne, dass sie die volle Personifizierung von Softwareagenten, Mensch-Computer-Assoziationen oder Multi-Agenten-Systemen nahelegen. Als Antwort auf die drei Risiken sollte vielmehr den Algorithmen jeweils ein auf ihre konkrete Rolle sorgfältig kalibrierter Rechtsstatus zuerkannt werden. Für das Autonomierisiko ist es eine adäquate Antwort, den Softwareagenten den Status als teilrechtsfähige Akteure zuzuerkennen. Ihre autonomen Entscheidungen sind mit Rechtsverbindlichkeit auszustatten und sollten Haftungsfolgen auslösen. Sie erhalten damit die beschränkte Rechtssubjektivität, als Stellvertreter bindende Verträge für andere abzuschließen. Zugleich sind sie in Fällen vertraglicher und außervertraglicher Haftung als rechtlich handlungsfähige Hilfspersonen anzuerkennen, sodass das maschinelle Fehlverhalten selbst (und nicht bloß das Verhalten der dahinterstehenden Unternehmen) einen Pflichtverstoß darstellt, für den die Unternehmen einstehen müssen.

world with the ulterior motive of creating a social structure that can control these risks to a certain extent in a preventive manner or at least be used for the settlement of damages.

This risk management concerns firstly the settlement of losses already incurred. In a situation of multiple causality, the law balances the losses in such a way that it creates an adequate financial pool which covers the losses and distributes the risk (“*deep pocket*”, “*risk spreading*”). Secondly – and perhaps more importantly – risk management means the collective management of future behavior. The law designates the limits of the risk pool in such a way that a realistic basis is created for the active and joint prevention of risks in areas where damage problems are concentrated. From both points of view, the law isolates the social space of collective responsibility in such a way that a functioning digital technology can develop to cope with the risks of digital networking.

## VI. RESULTS

The new digital risks – autonomy risk, association risk, network risk – confront private law with the challenge of redefining the legal status of autonomous software agents. Certainly not in the sense that they suggest the full personification of software agents, human-computer associations or multi-agent systems. Rather, in response to the three risks, the law should carefully calibrate the legal status of algorithms based on their concrete role. For the autonomy risk, it is an adequate answer to grant software agents the status of partial legal personhood. Their autonomous decisions should be legally binding and should give rise to liability consequences. This gives them the limited legal subjectivity to conclude binding contracts for others as a proxy. At the same time, in cases of contractual and non-contractual liability, they are to be recognized as legally capable assistants, so that the machine misconduct itself (and not merely the conduct of the companies behind it) constitutes a breach of duty for which the companies must be held responsible.

Während also eine klargeschnittene rechtsdogmatische Antwort auf das Autonomierisiko möglich ist, sind Lösungen für das Verbundrisiko und das Vernetzungsrisiko bisher nur in Konturen erkennbar. Eine denkbare Antwort auf das Verbundrisiko wäre ihr Rechtsstatus als Mitglied einer Mensch-Maschinen-Assoziation. Eine Maximallösung *de lege ferenda* würde den Verbund selbst als rechtliches Zurechnungspunkt für Handlungen, Rechte und Pflichten vorsehen – eine Lösung, die dogmatisches Neuland betreten würde. Eine Minimallösung *de lege lata* dagegen würde den Rechtsbegriff des Verbundzweckes einführen, der rechtsgeschäftliche Erklärungen der Softwareagenten ebenso wie ihre Rechte und Pflichten orientieren könnte. Die Antwort auf das Vernetzungsrisiko schließlich wäre der Status eines vom Haftungsrecht selbst autonom zu definierenden Risikopools, der die Rechtsstellung der Algorithmen im Kontext eines umfassenden digitalen Informationsflusses festlegen und die Haftung des Pools ausschließlich am Ergebnis eines rechtswidrigen Verhaltens des Pools festmachen würde.

Akteur mit beschränkter Rechtssubjektivität, Mitglied eines Mensch-Maschinen-Verbunds, Teilelement eines Risikopools – dies also sind die neuen Formen eines digitalen Rechtsstatus, deren Konkretisierung darauf auszurichten ist, ob und wie damit die eingangs geschilderten Verantwortungsdefizite bewältigt werden können. Aber zugleich muss ein solcher Rechtsstatus von Softwareagenten es ermöglichen, die Dogmatik der Rechtsgeschäftslehre und des Haftungsrechts konsistent und ohne unhaltbare Fiktionen weiterzuentwickeln, dies freilich nicht als Selbstzweck, sondern um im Falle von Schadensfällen und anderen sozialen Konflikten, die vor das Recht gebracht werden, die Forderung, Gleiches gleich zu behandeln und Ungleiches ungleich, auch im digitalen Raum zu erfüllen.

While a clear-cut legal response to the autonomy risk is possible, solutions for the association risk and the network risk are currently only visible in contours. A conceivable response to the association risk would be to give software agents the legal status as a member of a human-machine association. A maximum solution *de lege ferenda* would attribute actions, rights and obligations to the association itself – a solution that would break new ground in private law. A minimal solution *de lege lata*, on the other hand, would introduce the legal concept of association purpose, which could guide the interpretation of computer declarations as well as the determination of the participants' rights and obligations. Finally, the answer to the network risk would be a risk pool to be defined autonomously by the liability law itself, which would determine the legal status of the algorithms in the context of a comprehensive digital information flow and would establish the liability of the pool solely on the basis of unlawful behavior of the pool.

Actors with limited legal subjectivity, members of a human-machine association, elements of a risk pool – these are the new forms of a digital legal status. Their conceptualization must be geared to whether and how the above-mentioned deficiencies in responsibility can be overcome. But at the same time, any such a legal status for software agents must make it possible to further develop the doctrine of contract law and liability law consistently and without unacceptable fictions – not as an end in itself, but as the law's obligation to treat equal what is equal and to treat unequal what is unequal in the digital space as well.