Transformations of Responsibility in the Age of Automation: Being Answerable to Human and Non-Human Others

Mark Coeckelbergh and Janina Loh

Introduction

Rapid progress in automation, especially in robotics and AI, poses challenges due to its potential transformative power with regard to competences that were traditionally reserved for human agents. It has been suggested that formerly exclusive concepts—such as autonomy, agency and responsibility—might one day pertain to artificial systems in a similar fashion. Others argue that these concepts are only applicable to humans. The new technologies thus raise questions with regard to the meaning of these concepts.

The concept of responsibility is of high moral and legal status and plays a prominent role in every sphere of human acting. Every dimension knows its specific type of responsibility, due to the norms of identifying the responsible parties involved—e.g. moral, legal, political, economic, social responsibility, and various other forms of responsibility. Responsibility is an important tool for systematizing, organizing, and thereby clarifying intransparent and very complex situations that confuse the agents in question; situations, where classical ascriptions of duties and guilt frequently fall short. Unbundled properly, it can make sense of challenging hierarchical set-ups, an unclear number of involved parties, huge time-space-dimensions, and is able to complement traditional concepts like the concept of duties (Heidbrink et al. 2017).

But what does responsibility mean in the age of automation?

M. Coeckelbergh (✉) · J. Loh
Institut für Philosophie, Universität Wien, Vienna, Austria
e-mail: mark.coeckelbergh@univie.ac.at

J. Loh
e-mail: janina.loh@univie.ac.at

B. Beck and M. Kühler (eds.), *Technology, Anthropology, and Dimensions of Responsibility*, Techno:Phil – Aktuelle Herausforderungen der Technikphilosophie 1,
https://doi.org/10.1007/978-3-476-04896-7_2

Especially in the realm of technology, big data, and new media it is questionable if our traditional understanding of responsibility is able to face current challenges—mostly due to its restricted focus on the autonomous, self-sufficient, individual, human being as the genuine responsible agent. In the case of autonomous driving systems, for example, this distinguished focus on the single responsible person must be questioned: We can prospectively ask who should make the relevant decisions in certain circumstances on the street and how to distribute responsibility accordingly between the current occupants of the vehicle, its owner, the engineers/developers of the car, and maybe even the artificial system that is operating the car itself (Coeckelbergh 2016; Lin 2015; Loh and Loh 2017). In the next section we will differentiate between causal and moral responsibility and explain, why we focus on the latter. In conventional situations we identify the responsible agent via a set of competences that we commonly assume her or him to be equipped with—such as the ability to communicate, judgement, and autonomy (see the following paragraph). Given that autonomy is one of the crucial conditions for ascribing responsibility and assuming that the self-driving car is autonomous in a (morally, legally, socially, etc.) relevant sense, should it be trusted with making these decisions (Floridi and Sanders 2004; Wallach and Allen 2009; Misselhorn 2013)? If so, what would be the criteria to ascribe responsibility to it? Should and could an artificial system really be called responsible or even make a (morally, legally, politically, etc.) wrong decision? How would we go about confronting it with its responsibility in such a case? Even if there might be reasons to say that an artificial system cannot be responsible, it is worth discussing why (not).

Further challenges of the traditional concept of responsibility in the realm of technology and automation appear especially against a global backdrop—such as ascribing responsibility within the virtual sphere in general and within the sphere of social media or the global financial market in particular. In all of these cases, we cannot reduce responsibility to a limited and clearly defined group of responsible persons: due to e.g. intransparent contexts, the prominent involvement of algorithms, as well as the implicit redefinition of norms and the confusion of classical concepts such as privacy, personal identity, authorship, autonomy, and property (Floridi 2016). Again, our conventional methods of identifying individual human agents as the solely feasible responsible agents—or as other important functions within the relational setup of the traditional concept of responsibility such as the addressee and the authority (see the following paragraph)—frequently fail.

In this paper we argue that we need to question and move beyond a traditional understanding of responsibility (as outlined in the following paragraph) in order to update it for these and further challenges.

The Traditional Concept of Responsibility

In this paragraph we outline the traditional understanding of responsibility as it evolved since the term adjectivally firstly appeared in the thirteenth century in France (McKeon 1957; Bayertz 1995; Sombetzki 2014). This conventional

conception of responsibility rests on the ascription of competences as properties—its fundamental problem as we will elaborate on in the next paragraph. In order to analyze the basic structure of the responsibility concept, we articulate an etymological minimal definition and the five relational elements it includes as well as on the necessary conditions that are to be met for ascribing responsibility to someone.

Let us start with the minimal definition: A detailed etymological study (Sombetzki 2014, pp. 33–41) would show that "responsibility" means—firstly—"to be answerable for something". It is the ability to answer when someone needs to explain her- or himself (Coeckelbergh 2010; Duff 1998, p. 290; Heidbrink 2017; Kallen 1942, p. 351; Piepmeier 1995, p. 87). Secondly, responsibility is a normative concept, i.e. it is not only descriptive and causal. In—on the one hand—calling the sun responsible for melting the candle wax we use the term "responsible" in a metaphorical sense because the sun is not able to explain itself. In—on the other hand—calling someone responsible for killing another person we usually do not want to state a simple fact or see the person in question as a cause in a purely descriptive way. We want the claimed murderer to explain her- or himself and to accept her or his being guilty (Lenk and Maring 1992, p. 85; Werner 2006, p. 542). Finally, responsibility includes a specific psycho-motivational constitution of the responsible subject in question: we think her to be answerable in the sense of being an autonomous person, to feel addressed to take up her responsibility and to be equipped with several capabilities such as judgment and reflective faculty (Sombetzki 2014, pp. 39–41; Loh 2017).

This etymological minimal definition of responsibility leads to five relational elements. First, there is the individual or collective subject or bearer of responsibility as the responsible agent or person (the *who* is responsible?; Weischedel 1972). The subject is prospectively or retrospectively responsible for an object or matter (the *what* is x responsible *for*?). The subject is responsible to a private or official authority (the *to whom* is x responsible?; Ropohl 1994, p. 113; Schwartländer 1974, p. 1586) and *towards* a private or official addressee or receiver (Lenk and Maring 2007, p. 570). The addressee is the reason for speaking of responsibility in the context in question. We believe that this relatum is the most underestimated element within the relational setup of responsibility. Within the process of transforming this traditional understanding of responsibility we will suggest a radically different status for the addressee (as we will outline further below). Finally, the (private or official) normative criteria define the *conditions under which* x is responsible (Bayertz 1995, p. 13; Bierhoff 1995, p. 236; Forschner 1989, p. 591; Ingarden 1970, pp. 35–51; Lenk and Maring 2007, p. 570; Ropohl 1994, p. 113). They restrict the area of responsible acting and by this differentiate moral, political, legal, economic and other responsibilities, or better: domains of responsibility. A thief (= individual subject), e.g., is responsible for a stolen book (= retrospective object; better: the theft, a collection of actions that already happened) to the judge (= official authority) towards the owner of the book (= official addressee) under the conditions of the criminal code (= normative criteria that define a legal or criminal responsibility) (Sombetzki 2014).

What are the conditions for calling someone responsible? These are usually defined in terms of properties: In the light of this minimal definition of responsibility it becomes clear that a complex cluster of capacities is needed to call someone responsible. This set of properties includes 1) the ability to communicate (Piepmeier 1995, p. 86; Weischedel 1972, p. 15). The responsible agent needs 2) to be able to act, i.e. possess a demanding form of autonomy (Langbehn 2017; Lenk and Maring 1992, p. 77; Nida-Rümelin 2007, p. 60; 1998, p. 31; Schälike 2017). That includes (2.1) being aware of the consequences (knowledge), (2.2) being aware of the context (historicity), (2.3) personhood and (2.4) a scope of influence. Finally, for calling someone responsible it is necessary (3) that she or he is able to judge. This competence includes (3.1) several cognitive capacities such as reflection and rationality (Nida-Rümelin 2007, p. 71; Williams 2017) and (3.2) interpersonal institutions such as promise, trust, and reliability on the other (Bernasconi 2006, p. 224; Ricœur 2005, p. 357; Sombetzki 2014, pp. 43–62).

It is important to take into consideration that these three sets of capacities (communication, autonomy, and judgment) and the competences that come with them, can and should be ascribed in a gradual manner. As it is possible to speak of more or less communication skills, to say that someone is more or less able to act in a specific situation, she is more or less autonomous, reasonable, and so on, it follows that responsibility itself must be attributed gradually according to the present prerequisites. Assigning responsibility is not a question of "all or nothing" but one of degrees (Nida-Rümelin 2007, p. 63; Wallace 1994, p. 157).

If and only if we can reasonably assume the person in question to be equipped with this set of properties we are able to identify her or him to bear responsibility. The same (i.e. identification via a specific set of properties) holds for the other relata—especially the authority and the addressee (the thoughts outlined in this section are also to be found in a fairly similar way in Loh and Loh 2017).

Limits of the Traditional Understanding of Responsibility

As already claimed in the introductory paragraph, the age of automation confronts our traditional understanding of responsibility with several challenges, a discussion of which reveals the limits of the traditional conception. In the following we will have a closer look at these problems. They mostly appear due to its strict focus on specific sets of competences as properties that are seen as constitutive for its structure: subject, object, authority, addressee, normative criteria. Let us discuss the problems with each of these elements:

1. Identifying a responsible agent: The responsible person (i.e. the subject or bearer of responsibility, the answer to the question *who* is responsible?) is the core of the relational structure of our traditional understanding of responsibility. Whenever we identify someone to be equipped with the ability to communicate, with autonomy, and judgement, this person qualifies as a potential

responsible person. Responsibility might not be a sufficient characteristic for agency and personhood (we do not want to give a statement here because in this project we do not focus on agency but on responsibility). Whenever the need to ascribe responsibility appears, it is first and foremost imperative to find a responsible subject to bear this responsibility: no responsibility without an agent and no agent without the ability to bear responsibility. But in the age of automation, it becomes harder to ascribe responsibility due to the appearance of new potential responsible subjects such as machines, hybrid systems, and algorithms that do not exhibit the necessary properties while in specific contexts (such as the example of autonomous driving systems as sketched in the introductory part) almost intuitively being identified as responsible agents. Whether this intuition is justified, of course, requires verification. We think that the core structure of responsibility should somehow include these subjects of the age of automation—artificial systems and algorithms—in responsible relations. (In the next section we will propose solutions.)

2. Gaps in the relational setup of a specific responsibility: Although the responsible agent is the first and most important premise in ascribing responsibility during the pre-automation epochs the other relational elements are as well necessary for fully defining responsibility in a specific context. Without someone to whom the agent in question is responsible (i.e. the authority), it is not reasonable to speak of responsibility. Without normative criteria that define the rules and framework of calling someone responsible the responsible person cannot be judged and does not know how to act in the situation in question. Without an object or matter of responsibility, the subject of responsibility in question cannot know what its actions actually refer to. Finally, without an addressee one does not know why she or he is called to take up responsibility in the first place, why we speak of responsibility in this context at all. Without an addressee this responsibility ascription lacks the reason of its existence.

   A few words on the role of the addressee of responsibility: The addressee is situated at the opposite of the responsibility relation. She or he is the person affected by the responsibility in question and therefore defines the reason for its existence. An attribution of responsibility, that is to say, what the bearer of the said responsibility is called upon to do, changes according to its justification in relation to the person concerned by this responsibility. Imagine a thief (subject, individual) who has to answer for a stolen book (object, particular and retrospective) before a court (authority). Why does this whole procedure happen and for whom does it matter that the thief takes her or his responsibility? It seems that the person who has had the book stolen is the reason why the thief must answer for her or his actions, because she or he is affected by the theft. If she or he did not exist or if theft was not a crime in general, this responsibility would not exist. Moreover, the citizens of the country in question may be cited as secondary addressees of the criminal responsibility of the thief, or the norm

itself, because the thief has violated a law that is in the best interests of the citizens. He is primarily answerable for the stolen goods and secondarily for the citizens. They are the reason for his responsibility ("Where there's no plaintiff, there's no judge"; cf. Sombetzki 2014, pp. 113–118).

The five relational elements of the traditional concept of responsibility (subject, object, authority, addressee, and normative criteria) are completely defined via properties. Therefore, we are not able to handle situations in which e.g. we can already identify a responsible agent and the object of her or his responsibility but in which we do not have someone to whom the subject is responsible (i.e. the authority). In situations like these—when we are sure that someone needs to be called responsible for someone or something but do not know how to fully define one or all of the other relational elements—we are simply not able to adequately ascribe responsibility in its conventional understanding.

In order to address these problems, we want to transform this structural setup of responsibility with its five relational elements. This will enable us to more adequately handle situations in the age of automation when we are confronted with gaps regarding one or more of the relata.

3. Wide space-time-dimensions: The traditional concept of responsibility was initially "made for" moderately middle wide ranges of human acting; it became popular in the age of the first territorial and nation states and the Industrial Revolution when human acting became mediated via machines and instances, when the results of their actions became more incalculable, and the conditions of their acting became opaque. Whereas the ascription of duties and guilt structure the sphere of human acting that is hers or his closest realm (family, friends, her or his own live, profession, etc.) the concept of responsibility includes both—duties in the form of prospective responsibility (of course prospective responsibility and duties are not synonymous; cf. Sombetzki 2014, pp. 119–122), guilt in the form of retrospective responsibility—and reaches wider than these ancient types of modeling the close-realm human behavior. Now, in the age of automation we have reached the global sphere. Yet our conventional understanding of responsibility does not fit this wide horizon of human acting. For example, global political and economic crises such as climate change and the refugee crisis cannot be reduced to a limited and clearly defined group of responsible agents. And this is also and especially the case when it comes to technological action. There are global problems of ascribing responsibility adequately due to the rise of the age of automation: With the rise of digital technologies and in particular the internet, a new realm of human and non-human acting came into existence (or at least the technology radically transformed existing realms), social media open new perspectives of communication, cooperation, togetherness, and friendship, and the global financial market is primarily ruled by algorithms and parameters unseen by human eyes and often not completely understood by human minds. Again, our traditional understanding of responsibility, which primarily works in contexts within a moderate space-time-horizon, is inadedquate.

Transforming Responsibility

We see at least two strategies to overcome the limits of our traditional understanding of responsibility as outlined in the previous paragraph: A) moderately re-conceptualizing its structure in order to still hold on to its fundamental relational hierarchy and B) radically transforming this fundamental hierarchy of relational elements. While first sketching *both* classes of strategies to overcome the limits of the conventional concept of responsibility, we will eventually show that the first, moderate strategy does not offer satisfying solutions for the challenges of the age of automation. Ultimately, we will argue for the more radical strategy of transforming responsibility into a truly social-relational concept. Further below we will explain what we mean with "truly social-relational".

A) Moderately Re-Conceptualizing the Structure of the Traditional Concept of Responsibility

Regarding especially the problems 1) (identifying a responsible agent) and 2) (gaps in the relational setup of a specific responsibility) as outlined in the previous paragraph, one might be tempted to reevaluate type and number of the five relational elements (A.1). If it is true that we frequently fail to identify a responsible person and that in several situations in the age of automation further gaps in the ascription of responsibility appear regarding for instance defining an authority or a concrete object of responsibility, it might be easier and more efficient to delete the most difficult relational element or elements or at least to compensate one or more relational elements with new relata. For example, one may re-define the subject or bearer of responsibility (the *who* is responsible?) by focusing on collective forms of ascribing responsibility such as network, systems, hybrid, dynamic, shared, and distributed responsibility. Some approaches in fact seem to suggest to cancel the position of the responsible subject altogether in exclusively calling 'the system' responsible for something but no person or group of people whatsoever (Wilhelms 2017; Heidbrink 2012).

However, A.1) is still to be located within the horizon of the etymological minimal definition of responsibility. As outlined in the paragraph above, this minimal definition holds that responsibility means normatively and not purely descriptively the ability to be answerable and includes a specific psycho-motivational constitution of the responsible subject in question. This minimal definition determines the five relational elements in type and number. That means, from agreeing to the minimal definition follows the acceptance of these exact five relational elements (this thesis was explained in detail in Sombetzki 2014). Therefore, one cannot simply cut in this structural setup by deleting or re-defining one or more relata or by defining new ones—and especially not the relatum "subject or bearer of responsibility". If one agrees to the etymological minimal definition—and even those who

imply to cancel the subject of responsibility at least implicitly if not explicitly still cling to this traditional understanding of responsibility—one has to hold on to the relational elements that are included in this minimal definition.

The same is true for the idea of reconsidering and redefining type and number of the competences as conditions for calling someone responsible—the ability to communicate, autonomy, and judgement (see paragraph above) (A.2). Since the whole project of ascribing responsibility in the traditional manner rests prominently on the ascription of properties, we will always fail to concretely define responsibility in those cases when we are not able to assume that the entity in question is equipped with the necessary capacities and competences to be called for instance a responsible agent, a matter of responsible acting, an authority for judging the responsible person, an addressee or a normative criterion.

As long as we are not willing to reconsider the etymological minimal definition of the responsibility concept itself we will always be bound by the five relational elements and their necessary properties that follow from this etymological minimal definition (see paragraph above)! Every attempt to only moderately reevaluate type and number of the relational elements (A.1) or redefine type and number of the prerequisites for calling someone responsible or for defining the other relational elements (A.2) will at one point or another be internally inconsistent. We therefore suggest to radically transform the traditional understanding of responsibility: Radically transforming the basic structure of the conventional understanding of responsibility means—first and foremost—this: to reevaluate the three aspects of the etymological minimal definition of responsibility as outlined above. Responsibility is—firstly—the ability to answer when someone needs to explain her- or himself. Secondly, it is a normative concept that—finally—includes a specific psycho-motivational constitution of the responsible subject in question. To reconsider these three components does not mean to redefine the term "responsibility" itself. Besides, since we claim this minimal definition to be etymologically justified, it would not be feasible to redefine the term "responsibility" as long as we do not define a completely new term. Hence we hold on to the term responsibility, but propose to alter the basic structure of the traditional concept.

B) Radically Transforming the Basic Structure of the Traditional Concept of Responsibility

Due to the idea of verifiable properties at its core, the traditional concept of responsibility seems to mainly focus on the second and especially the third aspect of the minimal definition. This section of strategies instead increasingly shifts the perspective to the first part of the minimal definition—answerability—in order to outline a truly social-relational understanding of responsibility. Genuine or "substantial" or "strong" relationality is different from the simple or "thin" or "weak" relationality of the traditional concept of responsibility that only marks a linguistic status insofar as countless terms are dependent on a number of relational elements (e.g. "theft" is a relational term insofar as it requires a subject—the thief—and

an object—the stolen good). It *starts* from the idea that to be responsible is to be answerable to someone(s), and puts this at the *center* of the concept of responsibility, rather than seeing it as one of the elements of responsibility, as in the traditional conception of responsibility. What does that mean?

Let us first revisit the literature on answerability mentioned earlier in this paper in order to introduce and elaborate this different angle. In analytic philosophy, Watson (2004) has distinguished between responsibility as attributability (under which conditions can an act be attributed to an agent) and responsibility as interpersonal accountability or answerability, which concerns the conditions under which an agent can be asked to answer for what she has done by the members of the relevant moral community. With regard to criminal responsibility, Duff has argued that responsibility is to be understood in a relational way: 'to be responsible is to be held responsible for something by some other person or body within a social practice. To understand responsibility in these terms, we must answer three sets of questions: What is it to be a responsible subject? What are the proper objects of responsibility? To whom are we responsible?' (Duff 2005, p. 441). There are more relevant discussions about different senses of responsibility, for instance by Scanlon and by Shoemaker. What these accounts have in common is that they add an emphasis on relationality, and already give *some* "vertical" structure to the traditional concept of responsibility by prioritizing some questions rather than others. However, between these questions there is no hierarchy—they still put the different senses and questions regarding responsibility on the same level—and they start from the agent or person. Answerability is only *one* of the questions, and the agent remains central. These accounts thus remain relational in what we called a "thin" sense. A more "substantial" and arguably more radically relational conception of responsibility, however, does not start from the agent, person, or subject of responsibility, but from the *other*. This view can be articulated by drawing on the work of Emmanuel Levinas.

B.1) Levinas: The Other Is the Starting Point and Center of Responsibility Relations—Actors

For Levinas, responsibility starts not with the self but with the other. The other makes an appeal to me, asks or even demands a response. It is in the concrete experience of the encounter with the other, for instance in his case in the context of war, that the 'face' of the other (the way the other presents) makes a demand on me, which I cannot escape. It comes before freedom. Levinas writes: 'the face speaks to me and thereby invites me to a relation incommensurable with a power exercised' (Levinas 1991, p. 198) and 'the Other faces me and puts me in question and obliges me' (p. 207). In this view, then, responsibility is not about me but about the other. It is not so much about our willful and intentional acts, but about being asked to respond. Ethics for Levinas is not grounded in practical reason but is beyond reason (Bernstein 2002, p. 264) or at least reason comes in only after the response; first there is the face. The other and the relation to the other are central.

We have to transcend ourselves and see the other as unique and as other, without reducing to the same. The other is not an alter ego, another self. The other is also not an object which I can categorize. The other is the stranger.

Even if one does not accept Levinas's entire philosophy and view of ethics and responsibility, the very idea of an other-directed and other-centred ethics and conception of responsibility is and has been appealing. It means that relationality is not something that is so to speak stacked on top of the moral agent or subject of responsibility, but rather that the relation comes first, since according to this approach, the problem of responsibility arises from the other and my relation to the other—not from me. It is the other that invites me to a relation, and responsibility is then about my response to the other, within my relation and encounter with the other. Here the question is not about which properties I have that render me a subject of responsibility. Instead, I am called into responsibility by the other, and hence the other and the relation is first. This view radically transforms the structure of responsibility into a more vertical and asymmetrical one: other elements do not disappear but are part of language/reason (for Levinas both are the same) which come after the encounter with the other. Of course we can reason about conditions of responsibility. But in a substantial or strong relational conception of responsibility, what matters normatively speaking is the other: what the other asks, what the other needs, and our response to the other. Before my active response, there is a kind of "passivity" and "reception", in the sense that the responsibility is received, the response invited. This so to speak "takes over", "turns around", and transforms the entire structure of responsibility into a relation that starts with the other. The other takes a central and "highest" place in the structure.

B.2) Latour: Questioning the Human—Non-Human Hierarchy—Actants

Another more transformational intervention in the conception(s) and structure of responsibility is to question the exclusively human-centeredness of both the "weak" relational view (the traditional view) and the "strong" relational view articulated with Levinas. Here the idea is that whatever the precise structure of responsibility is, and whatever the precise relation between agent/person/subject/self and other is, we must question the assumption that the agent, person, subject, self, or other is necessarily a human being. Challenging this assumption raises at least the following questions: 1) Can the subject/agent/person of responsibility be a non-human? and 2) Can the other I am responsible and answerable *to* be a non-human? These questions open up the discussion to a potential intervention that does not change the structure of responsibility as such, but transform the concept by extending or expanding its scope to non-humans. Such an intervention may be inspired by arguments for including animals into the moral domain (Beauchamp and Frey 2011; Cohen 1986; DeGracia 1996; Donaldson and Kymlicka 2013; Nussbaum 2006; Regan 1986; Singer 1975) or by posthumanist and/or

non-modernist accounts such as Bruno Latour's (e.g. Latour 1993, 2004); Donna Haraway's (2000); Rosi Braidotti's (2013); Karen Barad's (2012, 2007) or Cary Wolfe's (2010) which suggest that things, machines, and cyborgs also deserve a moral and political place. For instance, we may ask if machines can be responsible (e.g. Coeckelbergh 2009; Floridi 2016; Loh 2016; Loh and Loh 2017; Wallach and Allen 2009), or if animals can have a 'face' and be 'others' that call us to respond (see for instance Coeckelbergh and Gunkel 2014).

While we are not sure if machines can be called "responsible" in any sense mentioned in this paper, it is worth elaborating the potential benefits of a more Latourian, posthumanist approach if that means that non-humans can at least be on the side of those we are answerable *to*. Indeed, combining the two interventions, we obtain a view of responsibility as answerability that is other-directed, whereby 'others' can be human or non-human. This solves the problems with the traditional conception in the following ways:

1. Following Levinas, there is no question who is responsible: *I* am responsible. Instead of trying to look for whom to blame, we should start with responding ourselves. For Latour, we are also responsible, without doubt: as representatives of things, we are responsible for their political representation. Who else could speak for them? They are mute, they cannot speak themselves. We have the responsibility of representation and translation. Beyond that, with Latour one could even argue that some non-human beings bear responsibility as well; however, we will not take this direction here but instead stay with our claim that the above-mentioned *human* representative responsibility is paramount.
2. It is clear who or what the addressee is: what matters is who/what appeals to my responsibility, asks a response, presents face. No scientific investigation is needed, at least in the first instance; what is needed is an encounter and phenomenon of 'face', of something/someone asking me to respond. According to Levinas, his demand and encounter is ethically prior. And if we follow Latour and interpret Levinas in a posthumanist way, it is regardless of their category 'human' or 'non-human'. Indeed, for Levinas the most unethical and irresponsible act is to categorize. This violence of categorization is avoided in a more posthumanist and non-modern account of responsibility, which is open to humans, non-humans, and hybrids (e.g. cyborgs). This can include animals (Coeckelbergh and Gunkel 2014). And Latour draws our attention to things, to which we are responsible as representatives. Again there is no question who the addressee is: through humans as representatives, things ask something, demand something, voice their "concerns". The whole paradigm of responsible relations and relata could include non-humans and hybrids. In this approach, there is a basic openness towards considering non-humans as addressees.
   That being said, embracing this approach does not exclude having ethical-philosophical arguments about *who or what* may reasonably be regarded as addresses or political discussion about *which* non-humans are the addresses, have face, and so on. Someone may object, for example, that rocks do not "demand" anything. And there may be no agreement in a particular society

about the moral status and "addressee" status of, say, an entity that has artificial intelligence. There is controversy about moral status, in philosophy and in society at large. But in the Levinasian approach, this discussion is not prior to the encounter and phenomenon but follows it. We can and probably have to discuss this, as philosophers but also as societies. But first there is the moral experience and relation. For example, it may be that there is *no* ethical encounter with a (particular) rock. Moreover, in the Latourian picture humans are still the representatives, they keep an important role: *they* can voice concerns, can be asked, and so on. And they do so as humans. In this way, an element of compatibility with the traditional, human-centred notion, can be retained.

3. In a global context several 'others' may present themselves to us, show their 'face'. We admit that it is not clear, based on the sketched account, how we (whoever that "we" includes) can "respond" in a global context, since it is a limitation of *both* the traditional conception of responsibility and the alternatives sketched here that they are focused on the person and do not easily scale to global and collective levels. Levinas's account is based on a very concrete face-to-face encounter. Perhaps in a global context the face-to-face ethics must be accompanied by other kinds of ethics and models of responsibility. Some may think about collective action and responsibility; others may argue that there is still a place for more abstract ethics à la Kant, for instance, but then it remains to be seen how such a different ethics and account of responsibility can work together with a more particularist and situationist approach like Levinas's. More importantly in the light of the questions raised in this paper, it remains to be discussed what the role of technology is viz-à-viz global responsibility. For example, it must be discussed if 'face' can show itself in ways that are mediated by technology. Moreover, we need to know more about how exactly the moral and political representation of humans and non-humans is supposed to work at a global level, and (again) how this may be mediated by technology. However, these questions open up an exciting research program, rather than undermining the proposed approach as such.

Conclusion

In this paper we have argued that for thinking about responsibility in the age of automation, the traditional conception of responsibility is insufficient. In response we have discussed some modest ("weak") proposals to revise that conception, but drawing on Levinas and Latour, we have also explored two more radical interventions which put the emphasis on the non-human other, in particular the non-human other as the one we are answerable to. For questions regarding automation, this implies that some technologies can either be seen as mediating our responsibility to the other or may themselves be considered and treated as addressee of our responsibility, in the form of things and other non-humans as well as hybrid beings that need to be represented. For our responsibility practices, these two

reconceptualisations of the relation between technology and responsibility require a reevaluation and possibly revision of our present ways of ascribing responsibility in various domains and contexts (local and global) where technology, and in particular automation, plays a crucial role.

With regard to a philosophical theory of responsibility, we also believe that more work is needed to further elaborate and, if so desired, defend the "strong" versions proposed here. Our main aim here was not to defend but rather explore what it would mean to take a Latourian or Levinasian approach, in response to problems with the traditional one. Our sketch of how this could pan out needs more work and further discussion. There remains a lot of tension with more traditional approaches. For example, those defending a traditional approach to responsibility (and moral status) may question giving addressee status to things (in general) or demand criteria for deciding which non-humans must be considered as addressees. But a Latourian approach questions this restriction to humans based on a non-modern approach, and a Levinisian approach questions this very project of categorization. By saying that there can be first an encounter and experience and then philosophical discussion and political argumentation, we have suggested a direction for dealing with this controversy and for supporting the latter, strong approach, but this needs further development.

Furthermore, more elaboration is desirable regarding the global dimension of ascribing responsibility where identifying responsible subjects is difficult and where notions such as the one we elaborated with the help of Levinas seem not to scale. More needs to be said about how notions of responsibility developed for local and personal contexts can be applied (or not) in global and network contexts in which humans and non-humans are involved. These paths of inquiry suggest a research program that is not only relevant to those who think about technology, but also more generally to everyone interested in the concept of responsibility, especially in how responsibility works in the real world with its technologies and global scope.

References

Barad, K. (2007). *Meeting the Universe Halfway. Quantum Physics and the Entanglement of Matter and Meaning*. Durham: Duke University Press.

Barad, K. (2012). *Agentieller Realismus. Über die Bedeutung materiell-diskursiver Praktiken*. Berlin: Suhrkamp.

Bayertz, K. (1995). Eine kurze Geschichte der Herkunft der Verantwortung. In K. Bayertz (Ed.), *Verantwortung. Prinzip oder Problem?* (pp. 3–71). Darmstadt: Wissenschaftliche Buchgesellschaft.

Beauchamp, T. L., & Frey, R. G. (Eds.). (2011). *The Oxford handbook of animal ethics*. New York: Oxford University Press.

Bernasconi, R. (2006). Von wem und wofür? Zurechenbare Verantwortlichkeit und die Erfindung der ministeriellen, hyperbolischen und unendlichen Verantwortung. In L. Heidbrink & A. Hirsch (Eds.), *Verantwortung in der Zivilgesellschaft. Zur Konjunktur eines widersprüchlichen Prinzips* (pp. 221–246). Frankfurt a. M.: Campus.

Bernstein, R. J. (2002). Evil and the Temptation of Theodicy. In S. Critchley & R. Bernasconi (Eds.), *The Cambridge companion to Levinas* (pp. 252–267). Cambridge: Cambridge University Press.

Bierhoff, H. W. (1995). Verantwortungsbereitschaft, Verantwortungsabwehr und Verantwortungszuschreibung. Sozialpsychologische Perspektiven. In K. Bayertz (Ed.), *Verantwortung. Prinzip oder Problem?* (pp. 217–240). Darmstadt: Wissenschaftliche Buchgesellschaft.

Braidotti, R. (2013). *The Posthuman*. Cambridge: Polity.

Coeckelbergh, M. (2009). Virtual moral agency, virtual moral responsibility: On the moral significance of the appearance, perception, and performance of artificial agents. *AI and SOCIETY, 24*(2), 181–189.

Coeckelbergh, M. (2010). Criminals or patients? Towards a tragic conception of moral and legal responsibility. *Criminal Law and Philosophy, 4,* 233–244. https://doi.org/10.1007/s11572-010-9093-6.

Coeckelbergh, M. (2016). Responsibility and the moral phenomenology of using self-driving cars. *Applied Artificial Intelligence, 30,* 748–757. https://doi.org/10.1080/08839514.2016.1229759.

Coeckelbergh, M., & Gunkel, D. (2014). Facing animals: A relational, other-oriented approach to moral standing. *Journal of Agricultural and Environmental Ethics, 27*(5), 715–733.

Cohen, C. (1986). The case for the use of animals in biomedical research. *New England Journal of Medicine, 315*(14), 865–870.

DeGracia, D. (1996). *Taking animals seriously. Mental life and moral status*. New York: Cambridge University Press.

Donaldson, S., & Kymlicka, W. (2013). *Zoopolis. Eine politische Theorie der Tierrechte*. Berlin: Suhrkamp.

Duff, R. A. (1998). Responsibility. In E. Craig (Ed.), *Routledge encyclopedia of philosophy* (pp. 290–294). London: Routledge.

Duff, R. A. (2005). Who is responsible, for what, to whom? *Ohio State Journal of Criminal Law, 2,* 441–461.

Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 374. https://doi.org/10.1098/rsta.2016.0112.

Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines, 14,* 349–79.

Forschner, M. (1989). Verantwortung. In Görres-Gesellschaft (Ed.), *Staatslexikon: Recht, Wirtschaft, Gesellschaft. Band 5, Sozialindikatoren – Zwingli* (pp. 589–593). Herder: Freiburg im Breisgau.

Haraway, D. (2000). A cyborg manifesto: Science, technology and socialist-feminism in the late twentieth century. In D. Bell & B. M. Kennedy (Eds.), *The cybercultures reader* (pp. 291–324). London: Routledge.

Heidbrink, L. (2012). Unternehmen als politische Akteure Eine Ortsbestimmung zwischen Ordnungsverantwortung und Systemverantwortung. *Ordo – Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft, 63,* 203–232.

Heidbrink, L. (2017). Definitionen und Voraussetzungen der Verantwortung. In L. Heidbrink, C. Langbehn, & J. Loh (Eds.), *Handbuch Verantwortung* (pp. 3–33). Wiesbaden: Springer VS.

Heidbrink, L., Langbehn, C., & Loh, J. (Eds.). (2017). *Handbuch Verantwortung*. Wiesbaden: Springer VS.

Ingarden, R. (1970). *Über die Verantwortung. Ihre ontischen Fundamente*. Stuttgart: Philipp Reclam junior.

Kallen, H. M. (1942). *Responsibility. Ethics, 52*(3), 350–376.

Langbehn, C. (2017). Identität, Autonomie und Verantwortung. In L. Heidbrink, C. Langbehn, & J. Loh (Eds.), *Handbuch Verantwortung* (pp. 315–336). Wiesbaden: Springer VS.

Latour, B. (1993). *We have never been modern*. (trans: Catherine, P.). Cambridge: Harvard University Press.
Latour, B. (2004). *Politics of nature: How to bring the sciences into democracy*. (trans. Catherine, P.). Cambridge: Harvard University Press.
Lenk, H., & Maring, M. (1992). Deskriptive und normative Zuschreibung von Verantwortung. In H. Lenk (Ed.), *Zwischen Wissenschaft und Ethik* (pp. 76–100). Frankfurt a. M.: Suhrkamp.
Lenk, H., & Maring, M. (2007). Verantwortung. In J. Ritter (Ed.), *Historisches Wörterbuch der Philosophie* (Vol. 11, pp. 566–575). Basel: Schwabe.
Levinas, E. (1991). *Totality and infinity: An essay on exteriority*. (trans: Lingis, A.). Dordrecht: Kluwer.
Lin, P. (2015). Why ethics matters for autonomous cars. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte* (pp. 69–85). Berlin: Springer Vieweg.
Loh, J. (2016). Verantwortung und Roboterethik – ein kleiner Überblick. *Humboldt Forum Recht, 03*(2016), 10–30.
Loh, J. (2017). Strukturen und Relata der Verantwortung. In L. Heidbrink, C. Langbehn, & J. Loh (Eds.), *Handbuch Verantwortung* (pp. 35–56). Wiesbaden: Springer VS.
Loh, J., & Loh, W. (2017). Autonomy and responsibility in hybrid systems – The example of autonomous cars. In P. Lin, K. Abney, & R. Jenkins (Eds.), *Robot Ethics 2.0*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780190652951.003.0003.
McKeon, R. (1957). The development and the significance of the concept of responsibility. *Revue Internationale De Philosophie, 6*(39), 3–32.
Misselhorn, C. (2013). Robots as Moral Agents. In F. Rövekamp & F. Bosse (Eds.), *Ethics in science and society: German and Japanese views* (pp. 30–42). München: Iudicium.
Nida-Rümelin, J. (1998). Über den Respekt vor der Eigenverantwortung des anderen. Interview mit Julian Nida-Rümelin, Professor für Philosophie an der Universität Göttingen. In B. Neubauer (Ed.), *Eigenverantwortung. Positionen und Perspektiven* (pp. 31–41). Waake: Licet Verlag.
Nida-Rümelin, J. (2007). Politische Verantwortung. In L. Heidbrink & A. Hirsch (Eds.), *Staat ohne Verantwortung? Zum Wandel der Aufgaben von Staat und Politik* (pp. 55–85). Frankfurt a. M.: Campus.
Nussbaum, M. (2006). *Frontiers of justice. Disability, nationality, species membership*. Cambridge: Harvard University Press.
Piepmeier, R. (1995). Zum philosophischen Begriff der Verantwortung. In F. Hermanni & V. Steenblock (Eds.), *Philosophische Orientierungen. Festschrift zum 65. Geburtstag von Willi Oelmüller* (pp. 85–102). München: Fink.
Regan, T. (1986). *The case for animal rights*. 2004 edition, updated with new preface. Berkeley: University of California Press.
Ricœur, P. (2005). *Das Selbst als ein Anderer* (2nd ed.). München: Fink.
Ropohl, G. (1994). Das Risiko im Prinzip Verantwortung. *Ethik und Sozialwissenschaften. Streitforum für Erwägungskultur, 1*(5), 109–120.
Schälike, J. (2017). Verantwortung, Freiheit und Wille. In L. Heidbrink, C. Langbehn, & J. Loh (Eds.), *Handbuch Verantwortung* (pp. 277–249). Wiesbaden: Springer VS.
Schwartländer, J. (1974). Verantwortung. In H. Krings, H. M. Baumgartner, & C. Wild (Eds.), *Handbuch philosophischer Grundbegriffe* (Vol. 6, pp. 1577–1588)., Transzendenz – Zweck München: Kösel.
Singer, P. (1975). *Animal liberation*. New York: HarperCollins.
Sombetzki, J. (2014). *Verantwortung als Begriff, Fähigkeit, Aufgabe. Eine Drei-Ebenen-Analyse*. Wiesbaden: Springer VS.
Wallace, R. J. (1994). *Responsibility and the moral sentiments*. Cambridge: Harvard University Press.
Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. New York: Oxford University Press.

Watson, G. (2004). Two faces of responsibility. In *Agency and Answerability. Selected Essays* (pp. 260–288). Oxford: Clarendon Press.
Weischedel, W. (1972). *Das Wesen der Verantwortung. Ein Versuch*. Frankfurt a. M.: Vittorio Klostermann.
Werner, M. H. (2006). Verantwortung. In M. Düwell, C. Hübenthal, & M. H. Werner (Eds.), *Handbuch Ethik* (pp. 541–548). Stuttgart: Metzler.
Wilhelms, G. (2017). Systemverantwortung. In L. Heidbrink, C. Langbehn, & J. Loh (Eds.), *Handbuch Verantwortung* (pp. 501–524). Wiesbaden: Springer VS.
Williams, G. (2017). Rationalität, Urteil und Verantwortung. In L. Heidbrink, C. Langbehn, & J. Loh (Eds.), *Handbuch Verantwortung* (pp. 365–394). Wiesbaden: Springer VS.
Wolfe, C. (2010). *What is Posthumanism?*. Minneapolis: University of Minnesota Press.

Mark Coeckelbergh is Professor of Philosophy of Media and Technology at the University of Vienna and the former President of the Society for Philosophy and Technology. He is also member of advisory bodies for technology policy such as the High Level Expert Group on AI, European Commission, and the Austrian Council on Robotics and Artificial Intelligence. He is the author of numerous publications, including the books *Growing Moral Relations*, *Human Being @ Risk*, *Using Words and Things*, *New Romantic Cyborgs*, and *Moved by Machines*.

Janina Loh (née Sombetzki) is university assistant (Post-Doc) in the field of philosophy of technology and media at the University of Vienna. She wrote her dissertation on the issue of responsibility—Verantwortung als Begriff, Fähigkeit, Aufgabe. Eine Drei-Ebenen-Analyse (Springer 2014). She published the first German Introduction to Trans- and Posthumanism (Junius 2018) and is about to publish an Introduction to Robot Ethics (Suhrkamp 2019). She habilitates on the Critical-Posthumanist Elements in Hannah Arendt's Thinking and Work (working title). Research interests: trans- and posthumanism, responsibility research, Hannah Arendt, feminist philosophy of technology, ethics in the sciences, and robot ethics.