



Synergies in Innovation: Lessons Learnt from Innovation Ethics for Responsible Innovation

Michel Bourban¹  · Johan Rochel²

Received: 23 July 2019 / Accepted: 2 January 2020 / Published online: 22 January 2020
© The Author(s) 2020

Abstract

This paper draws on the emerging field of innovation ethics (IE) to complement the more established field of responsible innovation (RI) by focusing on key ethical issues raised by technological innovations. One key limitation of influential frameworks of RI is that they tend to neglect some key ethical issues raised by innovation, as well as major normative dimensions of the notion of responsibility. We explain how IE could enrich RI by stressing the more important role that ethical analysis should play in RI. We focus on two transversal issues of IE: the issue of redrawing conceptual boundaries, especially the topic of the artificialization of the world, and the issue of responsibility, especially the notion of total responsibility. We address these two issues from the thematic perspective of IE, thereby generating lessons learnt for RI. These two examples are taken as illustrations and blueprint of the dialogue that should take place between the two fields.

Keywords Innovation ethics · Responsible innovation · Artificialization of nature · Synthetic age · Total responsibility · Technology

1 Introduction

Responsible innovation (RI) has established itself over the last few years as a rising field of scholarship and policymaking. Theories of responsible innovation assess the criteria and procedures according to which the processes and outputs of innovation can be structured or be made responsibly. This article is built upon the hypothesis that the

✉ Michel Bourban
bourban@philsem.uni-kiel.de

Johan Rochel
johan.rochel@gmail.com

¹ Christian-Albrechts-Universität zu Kiel, Boshstraße 1, 24118 Kiel, Germany

² Faculty of Law, University of Zürich, Zürich, Switzerland

emerging field of innovation ethics (IE) could contribute to the development of RI. It is an attempt to clarify how IE could enrich RI by stressing the more important role that ethical analysis should play in RI. To put it in plain terms, we want to conceptualize the ethical dimension of RI and contribute to its further development.

In contrast to RI, IE is not yet an established field among researchers and policymakers. Here, we propose to understand IE as a new field in applied ethics that deals with normative questions raised by social, economic, political and technological innovations. Among the diverse types of innovation which IE might address, we focus on technological innovation, in line with the literature on RI that we wish to contribute to. As explained by von Schomberg and Blok (2019: 5–7), RI scholars mainly understand innovation as technological innovation, thereby following a “techno-economic paradigm”.¹ As we explain below, we propose to construct IE as being composed of several subfields dealing with specific technological innovations (computer and information ethics, big data ethics, the ethics of blockchain, de-extinction ethics, and so on). In these subfields, specific thematic issues are dealt with by highly specialized communities of scholars. We have found two key transversal issues in these numerous subfields that could exemplify how to enrich the literature on RI: the artificialization of the world and new forms of responsibility.

Our main claim that IE can enrich the growing literature on RI requires us to first briefly present the current state of research on RI and IE (Section 1). Secondly, we exemplify our approach by discussing two examples of contributions of IE to RI (Sections 2 and 3). We use the same structure to address these two examples. We address a transversal issue found within several subfields of IE and we draw lessons learnt from them for RI. The first example focuses on the issue of drawing and shifting normative boundaries as a consequence of innovation (Section 2). The second example focuses on the concept of responsibility, as used within IE (Section 3). We will show how responsibility is used within specific subfields of IE and argue that we could learn important lessons for responsibility used as predicate of innovation within RI.

Our main claim is not about the historical development of these two fields but rather about stimulating a dialogue by mapping issues and concepts relevant to both fields. The general structure of this contribution might be compared with the approach proposed by de Hoop et al. (2016), in investigating the “family resemblances” between RI and other fields of research (in this case: participatory design/constructive technology assessment, development studies and Actor-Network Theory (ANT) approach). De Hoop et al. (2016: 113) assess these related fields with the objectives to “deepen the theoretical basis of RI and provide pointers for attention in our RI analysis”.

Similarly, we propose to deepen the theoretical basis of RI on two different levels. On the first level, we want to show the relevance of ethics within the field of RI. As we shall see below, the distinct subfields of IE are part of a tradition of applied ethics. Like RI, they deal with human actions and decisions in the specific context of technological innovation. In that sense, ethics (in the case of IE) and responsibility (in the case of RI) share common interests in human actions and decisions in conditions of uncertainty about consequences of technological and societal changes. They deal with related

¹ As von Schomberg and Blok (2019: 7) highlight, “Even though emerging frameworks of RI now attempt to shift away from this techno-economic understanding of innovation, the concept of technological innovation remains to prevail throughout both the literature and policy documents”.

issues, while adopting different tools and perspectives. This article is an attempt to lay down a methodological path to make the most out of these linkages. In this context, we will highlight the strong normative implications of the concept of responsibility, in contrast with many RI advocates, “who appear to be more interested in investigating the ‘ingredients’ or ‘pillars’ of responsibility than the normative dimensions of it.” (Pellé and Reber 2015: 107).²

On the second level, our objective is to diversify the community of scholars working on RI issues. So far, Science and Technology Studies (STS) and Constructive Technology Assessment (CTA) traditions have led the debate on RI (Timmermans and Blok 2018). The “detour” by IE will bring distinct fields of applied philosophy into the dialogue on RI. Identifying synergies between these fields of scholarship can bring together distinct epistemic communities, thereby leading to a dialogue on methods, concepts and values.

2 Defining Responsible Innovation and Innovation Ethics

This section aims to lay down preliminary qualifications of RI and IE. It starts from a broad and (hopefully) consensual definition, prior to enlarging the scope of investigation and explaining the possibility of a dialogue between RI and IE.

2.1 Innovation as a Focus of RI

As found in the literature, RI can be characterized in different ways, for instance as an “ideal”, a “strategy”, a “discourse” or a “discipline”. For the purpose of this contribution, RI is defined here as a field of research, as a rising subdomain within the philosophy of technology (von Schomberg and Blok 2019). De Hoop et al. (2016) stress that RI is ultimately grounded in a pragmatic ambition to improve real innovation processes.

For the preliminary qualification of RI, three definitional elements are important. Firstly, in its most basic sense, innovation is about bringing novelty into the world, mainly, but not exclusively, by creating new marketable products and services. This definition of innovation as novelty reflects Blok’s identification of the original understanding of innovation (Blok *forthcoming*; see also Godin 2009). This first element means that various actors shall be described as innovators. Following van de Poel and Sand (2018), we define innovators as agents who are involved in, and shape, the innovation process and the resulting innovative products and services. This definition includes a variety of distinct professionals having specific functions in the different stages of innovation, such as scientists, engineers, marketers, entrepreneurs and investors (Ziegler 2015).

Secondly, novelty must be complemented by a dimension of impact on the external world. This dimension is important to keep a distinction between innovation and invention (Godin 2015). In order to become an innovation, the invention must be brought into the world, for instance as a marketable product. This definition might

² In a very similar way, one reviewer has stressed that “many established frameworks in RI (think of the AIRR framework by Owen et al. 2013) are designed to think through the process of innovation, with a stronger focus on the social political legitimization of innovation practices (democratization), rather than moral legitimization”. We wish here to further enhance the normative dimension of RI.

include non-technological innovation. Social and political innovations bring new norms, regulations and cognitive frames to social and political practices. Although they are not directly brought to the market, these other types of innovation fulfil both the novelty and the impact criteria. For the present contribution, we shall remain as close as possible to the current focus of RI and put a clear emphasis on technological innovation in form of products and services which might deploy an important impact by being brought to the market.

Thirdly, if the general object of RI is innovation, the focus of investigation might be either on the process or on the outcome of innovation (product or service). The influential definition proposed by von Schomberg entails these two elements: “Responsible Research and Innovation is a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)” (Von Schomberg 2012: 50).

In its current state, the field of RI tends to focus on innovation as a process of R&D, within public or private organizations (for a critical view, Blok and Lemmens 2015). Most well-known is the definition given in the Rome Declaration on Responsible Research and Innovation as “the on-going process of aligning research and innovation to the values, needs and expectations of society” (European Commission 2014). This process-focused approach has consequences for the normative questions raised (oversight and stewardship, control of the process and its stakeholders, reflections on targets and objectives) and the places where these questions are raised (companies, public research institutions, society in general). The framework proposed by Owen et al. (2013) and Stilgoe et al. (2013) focuses on the process of innovation and deals with these questions, for instance by proposing normative conditions for the process of innovation (such as anticipation, reflexivity, inclusiveness and responsiveness for Owen et al. 2013). Overall, the predicate “responsible” is seen as an “add-on” which is used to describe the process leading to new marketable products and services: “with the help of this extension, innovation processes will be better enabled to balance economic (profit), socio-cultural (people) and environmental (planet) interests” (Blok and Lemmens 2015: 20).

In the meantime, RI might also be addressed from the perspective of the products and services made possible by the R&D process (outcome). In that sense, van den Hoven (2013: 82) explains that “responsible innovation is an activity or process which may give rise to previously unknown designs pertaining either to the physical world (e.g., designs of buildings and infrastructure), the conceptual world (e.g., conceptual frameworks, mathematics, logic, theory, software), the institutional world (social and legal institutions, procedures, and organization) or combinations of these, which – when implemented – expand the set of relevant feasible options regarding solving a set of moral problems”. Products and services, i.e. their quality and capacity to solve problems, are the focus of RI. As explained by van de Poel and Sand (2018), new products might help solve or mitigate moral problems by creating new options which address trade-offs between moral values. Echoing this double focus of research, van de Poel and Sand highlight that innovators have a twofold responsibility: “first for the process of innovation and second for the products (‘innovations’), the result of such processes.” (van de Poel and Sand, § 1.1).

These three definitional elements are taken as components of a working definition of the field of RI: an investigation into the processes which bring novelty to the world in the form of marketable products and services. This working definition is meant to describe as closely as possible the current field of RI. This is necessary in order to establish a dialogue with IE which could bring something to the field of research as it exists now. It is of course important to highlight that criticisms have been raised against the underlying techno-economic paradigm of innovation (Bontems 2014; Blok and Lemmens 2015; Godin 2015). According to the critical assessment by Blok and Lemmens (2015: 31), “the analysis of the concept of innovation which is presupposed in current literature on responsible innovation shows that innovation is self-evidently seen as (1) technological innovation, (2) primarily perceived from an economic perspective, (3) inherently good, and (4) presupposes a symmetry between moral agents and moral addressees” (see also Timmermans and Blok 2018).

Important parts of this critical assessment of the RI field of scholarship can be linked to ethics and, more specifically, to the function of ethical reflections in RI. Quite clearly, it might be possible to describe the entire RI project as being an “ethical” project in the sense of aiming at realizing values broadly shared (de Hoop et al. 2016: 112). Because RI is mainly conceived as process-focused, it is not surprising that these general ethical reflections have taken the form of procedural prescriptions formulated for innovation processes. With respect to the output of these processes, the deployment of innovative products and technologies has very often been assessed in broadly consequentialist terms (in line with the general pragmatic ambition of RI). As explained above, we address this under-conceptualization of ethics within RI in two different ways. Firstly, we make a “detour” through IE and the substantial ethical reflections linked to the different topics addressed. In this respect, it is interesting to compare our approach with the ambition of value-sensitive approach (like value-sensitive design, e.g. van den Hoven 2013). Both share the ambition to highlight the relevance of more substantial ethical values to assess innovation processes and output (Timmermans and Blok 2018: 4.2.2). Our approach is complementary to this avenue. Secondly, thanks to this first point, we will be able to show how other types of ethical theories, such as virtue ethics, might be relevant to RI.

2.2 Innovation as a Focus of IE

For the sake of the present contribution, we propose to define IE in a bottom-up fashion. Assuming the definition of innovation previously outlined, we searched for ethical reflections on specific innovative technologies. The technologies we focus on fulfil the two criteria previously identified: the novelty criterion and the impact criterion, more specifically defined as an impact on the world through marketable products and services. During our investigations, we found a number of highly specialized subfields dealing with specific technologies. Our goal was to navigate through these subfields and highlight key common issues they deal with. These common issues represent the core of what we call IE. Blogs, experts’ interviews and companies use this concept, but there is no such established field of research. We find mentions of the idea in scientific contributions on “innovation ethics” (de Morais and Stückelberger 2014) and on the “ethics of innovation” (Fabian and Fabricant 2014;

Gremmen et al. 2019), but so far there is no systematic attempt to define IE as a field of research.

The conceptualization of IE as a set of specific common issues related to innovation represents our attempt to link RI to other more specific subfields dealing with thematic innovations. The object we create and call IE is an intermediary between RI and these subfields. It allows us to identify, conceptualize and make usable the common issues the majority of these subfields have to deal with. In that sense, the present argument does not rely upon IE being an established field of research, but only upon IE being a plausible common set of issues. The issues we want to focus upon (namely our two illustrations below) share a common ethical feature. They broadly relate to the traditional question raised by normative ethics, namely the formulation and justification of norms which should guide human behaviors and decisions. In that sense, we do not focus on what others have called a philosophy of innovation, which could also entail elements of reflections not directly linked to behaviors and decisions (Pavie 2018). Similarly, we do not directly address the ontology of innovation (Blok [forthcoming](#)).

Following that bottom-up approach, we must first identify the main subfields dealing with specific technological innovations fulfilling the two conditions of novelty and impact. The list below is based on an analysis of recent literature dealing with these technological innovations (Fig. 1). The list is not closed and further subfields which fulfil the conditions could also be included. We have voluntarily set aside classical fields of applied ethics such as medical or military ethics. Because they are established fields of research, it is possible to propose the same kind of argument we are trying to formulate without conceptualizing an intermediary such as IE. There are certainly interesting lessons that we can learn to formulate by bringing RI and military/medical ethics into the kind of dialogue we advocate, but there is no requirement to go through an intermediary such as IE. In the meantime, it is clear that specific subfields do not exist in a vacuum. They rely upon conceptual and normative issues already dealt with by established fields of research. For instance, the ethics of drone warfare takes up parts of military ethics to focus on unmanned aerial vehicles; the ethics of geoengineering arises out of climate ethics but focuses on distinctive issues raised by climate engineering technologies.

2.3 Making IE Fruitful for RI

With these two working definitions of RI and IE, we are now in a position to specify the objective of this contribution: highlight how IE could enrich the literature on RI. Our approach might be compared with a detour with heuristic benefits. We consider the set of issues gathered under the umbrella of IE before focusing again on RI and explaining what we have learned on the way. This detour is organized alongside two transversal issues which we claim are relevant for the diverse subfields of IE: “redrawing boundaries” and “responsibility”. These issues are important for every single subfield within IE. As such, they are the theoretical objects around which we develop our dialogue between IE and RI. We will first reconstruct each of these issues as addressed by the subfields of IE, before explaining the potential for the lessons learnt to be applied to RI.

This approach might be illustrated by a very simplified innovation sequence. This sequence shall only be used to map the current focuses of the different fields of research and highlight the heuristic detour we propose. In each step, we mention the main

<i>Computer and information ethics</i>	Ethical debates on computational sciences and algorithms (see e.g. Fabris 2018)
<i>Big data ethics</i>	Ethical debates on the collect, treatment and use of data (see e.g. Mittelstadt and Floridi 2015)
<i>The ethics of blockchain</i>	Ethical debates on technology using blockchain (such as cryptocurrency) (see Tate, Johnstone, Fielt 2017)
<i>De-extinction ethics</i>	Ethical debates on resurrecting species (“Jurassic/Pleistocene Park”) (see e.g. Cohen 2014)
<i>The ethics of geoengineering</i>	Ethical debates on intentionally manipulating the climate system (see e.g. Preston 2017)
<i>The ethics of drone warfare</i>	Ethical debates on unmanned aerials vehicles (see e.g. Strawser 2013)
<i>The ethics of cloning</i>	Ethical debates on reproducing human beings artificially (see e.g. Häyry 2018)
<i>The ethics of human enhancement</i>	Ethical debates on improving, correcting, repairing humans (see e.g. Giubilini and Sanyal 2015)
<i>The ethics of doping</i>	Ethical debates in the field of sport (see e.g. Bloodworth, McNamee 2017)
<i>Energy ethics</i>	Ethical debates on the generation, delivery, use and effects of energy (see e.g. Sovacool and Dworkin 2014)
<i>Nanoethics</i>	Ethical debates on nano-scale technology (see e.g. Gordijn and Cutter 2014)
<i>The ethics of automatized systems</i>	Ethical debates on automatized vehicles and transportation modes (see e.g. Brundage 2014)
<i>Roboethics</i>	Ethical debates on robots, cyborgs and the man-machine interactions (see e.g. Lin, Abney, Bekey 2014)
<i>The ethics of cryonics</i>	Ethical debates about the cryopreservation of dead individuals at ultra-low temperatures (see e.g. Minerva 2018).

Fig. 1 The literature on innovation ethics

questions addressed by RI and IE. The figure clearly over-simplifies the reality of innovation processes by highlighting a single sequence with starting and finishing points, and hides the complexity of a wide range of deeply entangled forms of innovation.

This sequence allows us to highlight our general objective from another perspective. Research on RI has mainly focused so far on the process of innovation, the responsibility of innovators and the output of innovation process (A and B). Our hypothesis is that the different “thematic ethical issues” (IE) and the “general questions about ethical theory” (IE) raised by technological innovation (B) should be made explicit and brought to RI. Elements related to C and D are indeed relevant, for instance as implications of choices made during the process of innovation. We try to set aside all the other ethical issues raised by subsequent steps of the sequence (e.g. distributive justice questions) (Fig. 2).

3 First Transversal Issue: Redrawing Boundaries

All subfields of IE deal with the impact of their specific innovation on the way the majority of members of a given society at a certain time perceive what is “given”, “normal”, “natural”, “human”, etc. Technological innovations impact how we individually and collectively draw these various boundaries. We therefore label the set of transversal issues around this question as the “redrawing boundaries” issue. Individually in exceptional cases, or cumulatively most of the time, innovations question dichotomies between humans and machines, matter and mind, science and technology, natural and artificial. Far from being stable realities, our perceptions of these boundaries reflect conceptual lines evolving according to the current state of technological innovation. A qualitatively less important innovation (say an instrument accelerating a

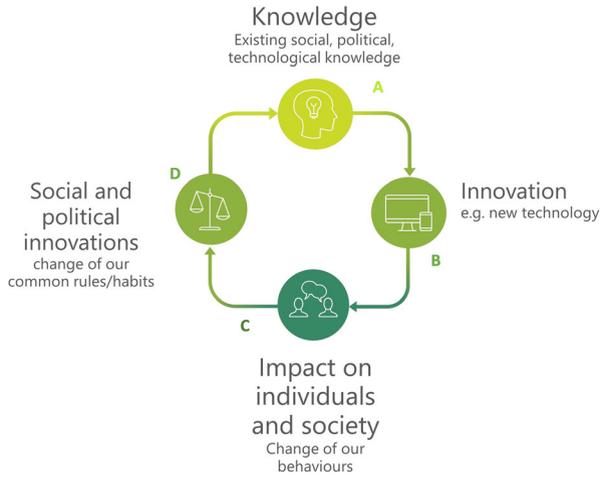


Fig. 2 Mapping innovation and its main ethical questions. (A) Current focuses on normative questions in the field of RI: responsible process of R&D leading to innovation, and responsibility of innovators. (B) Current focuses on normative questions in the field of RI: responsible products/outputs; in the field of IE: “thematic” ethical issues, and general questions about ethical theory. (C) Current focuses on normative questions both in RI and IE: impact of innovation, and distributive justice issues. (D) Current focuses on normative questions: Classical issues of political philosophy (e.g. legitimacy, justice, democracy), and ethics of transition: “dealing with losers” (Trebilcock 2015)

process) does not in itself deeply redraw a boundary. But its use and relevance for industrial purposes take place in the context of broader changes which impact these boundaries.

The relevance of these conceptual boundaries has been an ongoing topic of interest for empirically minded philosophers and psychologists trying to identify and measure the “framing effect” on behaviors and decisions (e.g. Frisch 1993; Kahneman and Tversky 2000). We pursue here a more normative ambition, as part of an applied ethics project. We focus on why conceptual boundaries which shift as a result of innovation are ethically relevant. This concerns the definition of conceptual boundaries and their use in public communication.

To illustrate the relevance of this issue and what we could learn from it for RI, we propose to focus on a recurrent topic in a number of subfields: the *artificialization of the world*. Other transversal issues would have been thinkable, for instance on the moving boundaries between humans and machines with robots and nanotechnologies. We choose the artificialization of the world for three reasons. First, the argument that current technological innovations redraw the line between the natural and the artificial is influential in most subfields in IE, especially in debates on artificial intelligence, nanotechnologies, geoengineering, de-extinction, cloning and human enhancement. In this sense, it is a truly transversal issue. Second, this topic allows us to link the literature on RI with new fields concerned with the impact of technological innovation on the natural environment, such as environmental ethics and climate ethics (for a similar approach, from the perspective of agricultural ethics, see Gremmen et al. 2019). This contributes to the core objective of this paper, that is, to identify synergies between distinct fields within the broad field of philosophy of technology and bring together distinct epistemic communities working so far mainly separately from each other.

The third reason is that this topic is in line with the main objective of the RI discourse, that is, to steer technological innovation in a responsible and desirable direction, especially with regard to the great societal challenges of our time, such as climate change (von Schomberg 2013). Assessing the impacts of technological innovations on the natural environment, especially of geoengineering projects whose objective is to counter global warming or offset some of its effects, is therefore directly relevant for scholars and policymakers interested in RI. A technophile view could consider the Synthetic Age as a possible way to remedy global and intergenerational ecological problems in the Anthropocene, for instance by considering geoengineering as a solution to climate change (Crutzen 2006) or green nanotechnologies as instruments to promote environmental sustainability (Shah et al. 2014). In that context, IE develops a more critical approach to these technological innovations, for instance through the ethics of geoengineering and nanoethics.

3.1 The Artificialization of the World

Most often, the artificialization of the world argument takes the form of an ethical objection against new technologies because of their effects on the natural environment. This objection has considerable appeal across different subfields of IE. We illustrate it here with the case of climate engineering. We briefly sketch the context in which the argument is used and what this use reveals for IE.

The objection of the artificialization of the world posits a clear distinction between the “natural” world and the impact human beings has on it, resulting in a part of nature becoming somehow changed into an “artefact”. This objection has its roots in ancient philosophy. According to Aristotle, while a natural object has a principle of movement and stationarity within itself, an artefact lacks the source of its own production; its principle lies in something external. This external source is the intentional action of a human: artefacts display the influence of human intention in a way that natural objects do not (Preston 2012: 191).

Contemporary scholars in environmental philosophy draw on this account and add a normative component to criticize current technologies (Lee 1999; Katz 2015). Artefacts are ontologically different from natural entities. They have a different essence, a different kind of being. While artefacts embody human intentionality, a truly authentic natural entity lacks any conception of design. An artefact has a different metaphysical status from a natural object. While nature is arguably intrinsically valuable, in the sense that it has a value independent from human purpose, activity and interest, an artefact’s value is instrumental to the human purpose used for creating it. Since the process of artefacting leads to a loss of intrinsic value, natural things are generally good and artefacts are generally bad. The natural is associated with the normal and the safe; the artificial is associated with the suspicious.

This objection is widely used in the case of climate engineering technologies, which “aim to deliberately alter the climate system in order to alleviate impacts of climate change” (Boucher et al. 2013: 627). There are two climate engineering methods: carbon dioxide removal (CDR) techniques, the purpose of which is to reduce the levels of already emitted carbon dioxide in the atmosphere, and solar radiation management (SRM) techniques, which aim to reduce solar radiation that reaches Earth. These methods are currently the object of numerous technological innovations. Marketable

products and processes are being developed across the globe, and entrepreneurs are registering patents for various techniques.

In line with the bulk of the literature on the ethics of climate engineering (see for instance Preston 2017), with focus here on solar geoengineering projects, which are correlative to the advent of the Synthetic Age (Preston 2018). So far, most of the major global impacts humans have had on the natural world have been inadvertent, such as modifying the chemical composition of the atmosphere, depleting the ozone layer, reducing the genetic diversity of plants and animals, acidifying the oceans and disrupting the nitrogen cycle. None of these phenomena marking the entry into the Anthropocene was caused intentionally. With ongoing technological innovations, we are in the process of replacing some of nature's most historically influential operations with synthetic ones of our design—how DNA is constructed, how ecosystems are composed, the amount of solar radiation reaching the Earth. From the atom to the atmosphere, new technologies have the potential to remake the natural world by replacing unplanned physical and biological operations by conscious, intended processes. Our planet is becoming increasingly malleable to us, something we design.

Environmental ethicists such as Katz (2015) object to SRM techniques that they involve the artificing of nature by changing some of the Earth's basic conditions to restore more stable climatic conditions for our societies. An artificed climate would involve human agency; it would embody human intentional structure. Engineering our climate would turn the Earth into a giant artefact. According to this objection, by changing the metaphysical status of the Earth, we diminish its value: nature is altered by human plans and the planet is reduced to a “zoo” or a “garden” that needs our management skills and technological means to flourish (Katz 2015: 494).

The artificialization of the world, as illustrated by solar geoengineering, is an example of an objection triggered by the fact that a conceptual boundary is moving. Innovation creates something new, and this novelty forces us, individually and collectively, to redefine what is “natural” and what is “artificial”, in this case the climate system. Lee and Katz view this artificialization process with a critical eye. So does McKibben, who claims that the artificialization of the world has led to the “end” or the “death” of nature. The independent, wild, pure nature has left the stage: all that remains is an artificialized nature, a world in which “each cubic yard of air, each square foot of soil, is stamped indelibly with our crude imprint” (McKibben 1989: 96).

While this critical reaction to the shift of boundaries between the natural and the artificial is helpful in understanding a core ethical concern raised by technological innovation, other scholars have developed a more fine-grained approach. The main weakness of the objection supported by Lee, Katz and McKibben is the too-sharp dualism between the human and the natural it assumes. This divide fails to acknowledge different degrees of naturalness and artificiality in objects (Preston 2012: 191–193; Cohen 2014: 167–168; Rolston III 2017: 64–65; Holy-Luczaj and Blok 2018). The artificial and the natural are the two opposite poles of a continuum, rather than two exclusive metaphysical statuses. Naturalness is a variable, ranging from dominantly natural to dominantly artificial. There is still plenty of nature in the Anthropocene, just like there has constantly been some observable anthropogenic influence on nature.

Assuming that there is some degree of naturalness in the artificial, we actually find a more accurate argument that highlights the ethical worries raised by some large-scale technological innovations. The argument focuses on the particular kind of uncertainty

we have to face in the Synthetic Age. There is an inevitable naturalness that remains in artefacts. Despite the best efforts of human designs, there will always remain a gap between the intentions of innovators and the actual behavior of their creations over time. If solar geoengineering technologies were deployed, the whole climate system would become more artificial, and the interaction between the artificial and the natural elements within the system would be particularly unpredictable and could be potentially catastrophic. Preston (2012: 196–197) explains that “the unpredictable effects of SRM are layered on top of the unpredictable effects of anthropogenic warming”. Compared with a very complex artefact like an airplane, for which much can be done to reduce unpredictable side-effects, “the climate system involves complexity of a different order – a type of ‘hyper-complexity’ – making the elimination of uncertainty much harder, if not impossible”. This deep uncertainty is a crucial reason why many philosophers oppose solar geoengineering projects. Resurrecting species, modifying DNA or introducing nanotechnologies can have similar, deeply uncertain side effects. Deep technologies that lead to profound transformations in reality and radical changes in the boundary between the natural and the artificial are characterized by deep uncertainties. This is not a sufficient reason for not pursuing these innovations, but it is certainly a good reason for pursuing them with great caution.

3.2 Contributions to the Literature on Responsible Innovation

The transversal issue of “redrawing boundaries” is interesting for reflections on RI in that it might help enriching preliminary definition of the kind of “novelty” which innovation brings into the world. We highlight two major contributions here of IE to RI.

Firstly, our reflections make clear that any definition of RI needs to integrate the fundamental relevance of conceptual boundaries. There are two levels of relevance. On a first level of relevance, innovation changes the way we conceive major boundaries that help us make sense of the world around us. The novelty brought into the world affects the very concepts that we use to make sense of the world. History or sociology of sciences and technology could measure and assess how these boundaries have shifted over time (for a classical piece on human and artificial intelligence, Woolgar 1985). Innovations redraw the content of the reference frame expressed by a boundary. Conceptualizing innovation must also be about trying to grasp the implications of novelty that changes the conceptual tools which we use to perceive the world.

The second level of relevance is more normative: the active use of these boundaries. Generally, concepts such as “nature” or “artefact” in our example, and more broadly “given”, “normal” and “human”, are by no means neutral from an ethical point of view. How much and how innovation impacts conceptual framings are questions with a normative dimension. As institutions or as individuals, the way we use these concepts (and thereby assume specific boundaries) is not ethically neutral. This point is independent from the specific argumentative use of these boundaries (as objection, as general framing, as value, etc.). Such concepts have strong normative components, which should be duly integrated as object of interest into RI. In that sense, RI should be about the processes and outputs of innovation, but also about the responsible use of conceptual framings.

The artificialization of the world is one example of an objection used across different subfields of IE and based upon a specific understanding of the boundaries at stake. The

point that “innovation arguably does (partly) transform natural matter into artefacts” might be criticizable if one defends a specific view of the value of nature. The general point for RI is here to acknowledge the requirement to shed light on underlying assumptions of how the poles of the various boundaries, such as the “natural” and the “artificial” poles of the continuum identified above, are ethically defined and justified. It helps investigate which assumptions are defended when it comes to draw upon poles such as “nature”, “human” or “normal”. As our short detour in the ethical debates on geoengineering has shown, there are distinct underlying positions on the value of nature. Our argument is not in favour of or against specific understandings of these poles, but only, because of our responsibility, as a requirement to be explicit as to which position one holds on them.

The sociological and normative levels of relevance are distinguished for analytical purposes, mainly in order to highlight a link to our individual and collective responsibility. This necessary discussion is not only theoretically relevant, but it has very practical implications for the focus on responsibility within RI. For instance, this “general boundaries” issue is linked to the question of the “burden of proof” by raising the question of who should be able to prove what. In the example discussed above, there seems to be a strong normative assumption in favour of the value of “nature”. In short, what is natural seems to be normatively desirable, whereas the artificial is commonly perceived as ethically suspicious. This means that the burden of proof in case of novelty is put upon innovators who challenge the putative meaning of “nature”. Someone changing natural conditions bears the burden of proof to make a case for his/her innovation.

Boundary issues are everywhere, but we should not see them as negative external constraints. On the contrary, they offer us entry points into the further conceptualization on the assumptions upon which innovation relies and their implications in terms of burden of proof. This responsibility to be as explicit as possible regarding their own assumptions on boundaries might be particularly strong for innovators, i.e. as defined above, all the actors actively involved in innovation processes. Following the mapping of types of responsibility for innovators proposed by van de Poel and Sand (2018, § 3.3), this specific type of responsibility might fulfil the functions of “maintaining moral community” and of “giving due care to others”. By being explicit and transparent about their definitions and use of boundaries, innovators promote public trust in innovation (“maintaining moral community”). This specific element could be part of a backward-looking understanding of responsibility as accountability. Innovators are accountable for their use of categories and boundaries. In the meantime, their effort to be as explicit as possible has a forward-looking dimension. Innovators show their capacity to take care of an open and honest public discourse on innovation, its conceptual framings and their consequences for society. As contribution to the general conditions of innovation, this responsibility might be understood as an instance of the kind of stewardship Stilgoe et al. call for, namely “taking care of the future through collective stewardship of science and innovation in the present” (Stilgoe et al. 2013: 1570). This point also connects to a view of responsibility as part of a broader set of virtues which innovators should display (Sand 2018).

The second contribution of this reflection on drawing boundaries across subfields of IE to the discourse on RI is to highlight the collective dimension of the debates. RI scholarship considers political decision-makers and the broader public as relevant

actors of the innovation process. For this reason, RI has an important tradition on arguing that responsible innovation processes should be the object of public debates. As listed by Owen et al. (2013), we need to address questions such as: What is the purpose of innovation? What kind of society do we want to create? What kind of future do we want innovation to bring into the world? Technological choices should not be left to the innovators alone. We need to have a democratic discussion about the kind of future we want in order to evaluate and control technological innovation.

We underline this conclusion by arguing from the perspective of IE. Boundaries issues should be at the core of public debates on innovation. They represent the conceptual tools used to formulate public debates and, in the meantime, they offer arguments in addressing specific technologies during these debates. This collective dimension highlighted from the perspective of IE creates an opportunity to make more out of the point made by Blok about the necessity to address power as a key element of the RI process (Blok 2018: 3, Blok and Lemmens 2015: 32). Power is defined as the capacity to impact the R&D process and its output, but also to impact the required deliberative space and its conceptual framings. Economic giants have the power to define what the notion of “human” should encompass and thereby can impact the public debates in a decisive way. If we want to keep the idea of a balanced public debate, we need to design institutional arrangements which are able to address imbalances of power in the framing of the technological challenges and in the range of decisions open to the public. IE fields might represent sources of inspiration for such mechanisms.

Let us mention two examples linked to the social and political components of IE. Firstly, scholars in environmental ethics argue that we need to develop a new concept of citizenship, an ecological citizenship based on a duty of global and intergenerational justice to ensure that one’s ecological footprint is sustainable. At the individual level, Dobson (2003: 119) stresses that “the ecological citizen will want to ensure that her or his ecological footprint does not compromise or foreclose the ability of others in present and future generations to pursue options important to them”. At the institutional level, Bourg and his colleagues (Bourg and Whiteside 2010; Bourg et al. 2017) have recently proposed a new model of democracy, an “ecological democracy” based on new forms of political deliberations between scientifically informed citizens able to adapt their economy and their society to current environmental issues such as climate change and biodiversity loss. The idea is not to give the power to the scientists (technocracy) but to reform representative democracies by developing new institutions and rethinking the role of the citizen to help democratic societies to face the challenges of the twenty-first century.

Secondly, the field of AI is marked by the search for suitable governance frameworks. These governance frameworks have the objective of setting up consultation mechanisms with all the relevant stakeholders. A mapping published by researchers in June 2019 has identified as many as 84 public and private initiatives (declarations, charters, ethical frameworks) on the ethical dimensions of AI.³ These initiatives are important in institutional settings where the drawing of boundaries occurs (through and within the initiatives). Interesting is the fact that the designers of these initiatives (experts, public servants, policymakers, companies) have the ambition of being self-reflective in dealing with these framings and boundaries. An essential part of these

³ Preprint of the paper entitled “Artificial Intelligence: the global landscape of ethics guidelines” (Jobin, Ienca, Vayena). Available on <https://arxiv.org/abs/1906.11668> (accessed 1st of July 2019)

initiatives is to address what “artificial” (in contrast to “human”) could and should mean. Of course, these initiatives face the challenge identified by Blok and Lemmens of bringing the relevant stakeholders to the table and addressing imbalances of power (Blok and Lemmens 2015). With respect to this challenge, these frameworks seem promising because they provide a balanced and self-reflective space for these debates. Power relations are not negated, but—in the best case—they are organized and channelled under the umbrella of public authorities taking responsibility to gather scientific experts, economic actors, public institutions and the broader public. The existence of these frameworks, their setting and the stakeholders involved increase transparency of the terms and boundaries at stake in the debate.

4 Second Transversal Issue: Responsibility

Innovation necessarily involves individual and collective human actions. Something new is created and brought into the world. Innovation is a process that involves human intentions. This does not mean that all implications are foreseen or foreseeable. However, innovation does not “happen” beyond or without any human action. This leads us to the key ethical issue of responsibility as dealt with in the distinct subfields of IE.

As for the “moving boundaries” issue, we want to focus on responsibility as a transversal issue across subfields of IE and ask what we could learn for RI. It means that, for the time being, we do not use “responsibility” as a predicate of innovation (either process- or product-focused) as in the literature on RI (van de Poel and Sand 2018). As proposed by van de Poel and Sand, we want to identify potential lessons learnt with relevance to RI.

We start with a brief presentation of new forms of responsibility in specific subfields. Second, we present the relevance of the notion of “total responsibility” in the Synthetic Age, especially in the case of geoengineering. Third, starting from this example, we highlight the different conceptions of “ethical theory” which are found in the approaches in different subfields to responsibility. Finally, we summarize how these concepts of responsibility might contribute to RI.

4.1 Accounting for New Forms of Responsibility

Responsibility is one of the most complex normative notions in modern and contemporary practical philosophy. Normative responsibility is a polysemic notion that can be both positive and negative, both causal and remedial, both backward-looking and forward-looking. Although a conceptual mapping of the different meanings of responsibility that are used in moral philosophy is beyond the scope of this paper (see Pellé and Reber 2015), it is interesting to note that classical conceptions of responsibility are increasingly complemented by new forms of responsibility that try to grasp the implications of current technological innovations.

Traditionally, normative responsibility has been attributed in cases where (Jamieson 2010):

- An individual acting intentionally harms another individual.

- Both the individuals and the harm are identifiable.
- It can be shown that the individuals and the harm are connected.

Certain features of innovation deeply challenge this classical conception of responsibility. Nanotechnologies, de-extinction technologies, geoenvironmental technologies, cloning technologies, human enhancement technologies and robotic technologies make clear that we need a refined version of responsibility. Several effects of these technologies are uncontrollable, unpredictable and irreversible. Some might deeply alter natural conditions; some might even modify our human condition. Most of these technologies result from a collective enterprise; a multi-actor process composed of individuals whose motives are varied and who generally do not intend to harm others. In such uncharted territory, any account of responsibility needs to include several complex elements (Frogneux 2015): the integration of collective agents (companies, states, international institutions), the integration of non-human victims or even non-existing victims (future generations) and complex causal chains (global and intergenerational).

These issues are not fundamentally new to specialized philosophical literature on responsibility. These different forms of “extension” to the concept of responsibility have already been addressed as theoretical challenges by scholars in moral and political philosophy. The more practical dimension of these extensions has been addressed by political philosophy regarding global and intergenerational justice (for overviews, see Lichtenberg 2010; Jamieson and Di Paola 2016). Theories of justice applied to global trade or climate change have also integrated complex accounts of responsibility (Risse 2017).

It is however interesting to note that these theoretical and practical issues are gaining new momentum in the context of the different sub-topics of IE. Every subfield is confronted with issues around responsibility, found in the context of a specific type of innovation. In looking for responsibility-issues across IE, we found the following cluster of questions, each of them echoing one of the elements previously mentioned:

- How to make sense of aggregative responsibility? The concept of responsibility should integrate small-scale innovations which cumulatively contribute to harming specific interests of individuals or communities, often unintentionally (e.g. a start-up inventing new ways to track people’s data online, knowingly or not, deliberately or not, participating in the general surveillance of citizens).
- Who or what exactly should be considered when assessing one’s responsibility? The concept of responsibility should be inclusive, i.e. able to make some normative room for interests which go beyond currently living human beings, in determining if any harm has been committed (including at least members of future generations, non-human animals, species, ecosystems).
- How to make sense of a complex chain of implications in situations of deep uncertainty? The concept of responsibility should allow accounting for both the number of actors involved in a causal chain (inventor, producer, raw material producer, designer, seller, consumer), and for its temporal dimension (an innovation might impact individuals over several years/generations).

4.2 Total Responsibility in the Synthetic Age

To exemplify what responsibility addressed from the perspective of IE could amount to, we focus on the case of “total responsibility”. The objective is not to assess the strengths and weaknesses of this specific idea but to use it as a way of illustrating potential lessons learnt from IE for RI. This illustration is also useful in addressing the three clusters of questions previously identified.

“Total responsibility” is a concept used in specific subfields of IE. With de-extinction technologies, nanotechnologies or geoengineering technologies, innovators must take responsibility for the new synthetic systems they choose to engineer. They have to assume interventionist technologies that bring human design deeply into processes used to give the natural world its shape. They have to take responsibility for the Synthetic Age by choosing which new or old species go extinct or get to survive, how to restructure matter at an atomic and molecular level or how to manipulate the climate system. With the technologies of the Synthetic Age, innovators have for the first time the power to rebuild nature as they see fit: “To take up this role deliberately and on a global scale is completely new territory for our species” (Preston 2018: 104).

This view is especially influential in the fields of geoengineering and human enhancement ethics, particularly because what is at stake in these fields (the human condition, human species, the climate, life) is perceived as absolutely fundamental. Crucially, these subfields are the same as those already addressed in the first transversal issue: where boundaries are substantially shifting, crucial debates on responsibility of a new kind emerge.

The notion of total responsibility correlates with existential anxiety concerning the fact that our plan to artificialize nature will fail, leading to unprecedented disruption (Preston 2012: 197–199). For instance, if solar geoengineering technologies were to be deployed, the researchers, innovators and people in charge of the implementation would not only be responsible for artefacting nature by deliberately altering the climate system; they would also be responsible for the potentially massive adverse side effects of this manipulation of the Earth system (for an overview of such effects, see Robock 2016). They would become responsible for the ecology of the Earth as a whole, for managing the entire planet. At every moment, geoengineers should be responsible for ensuring that the climate is hospitable for all. They should assume the potential damages they could impose on current and future generations, as well as other living beings.

This new responsibility for this kind of deep technological innovation has the three key characteristics identified above. First, it is aggregative, since the innovation results from the aggregation of the actions of multiple agents, such as the researchers at the origin of the geoengineering project, the scientific and economic partners who develop it, the people and states who implement it and the institutions and organizations that are responsible for governing the project and monitoring its effects. Second, it is inclusive, in the sense that side effects of any intentional manipulation of the climate system would impact both human and non-human forms of life on the planet (including ecosystems). Third, it is characterized by deep uncertainty, in that the unpredictable effects of geoengineering are layered on top of the unpredictable effects of climate change. Intentional manipulation of the climate system (geoengineering) takes place in

a context of unintentional change of the climate system (anthropogenic global warming). Since manipulation of ecosystems is already characterized by very uncertain side effects due to our structural ignorance of the complexities of the functioning of natural systems, the manipulation of a global natural system is a deeply uncertain project. It is especially this deep level of uncertainty at a global level that makes the responsibility of those involved in geoengineering “total”.

4.3 Ethical Theory

In this section, we argue that reflections on new forms of responsibility, illustrated here by “total responsibility”, bring to light an additional underlying issue: the crucial role played by the type of normative ethics one more or less implicitly assumes (consequentialism, deontological ethics, virtue ethics). Our objective is firstly to locate this issue as part of the debate and, secondly, to notice the recent rise of virtue ethics in debates on IE. These two elements can be illustrated by the objection of hubris.

In the context of “total responsibility”, the hubris objection typically takes the form of a “playing God” critique (Hartman 2017). It is used as an objection against specific innovations. For instance in the field of human enhancement or in geoengineering, this objection is an important part of the debate on responsibility. In the case of technological innovation, innovators show hubris if they overestimate their technical and epistemic abilities. This overestimation might come with an underestimation of the probability of failure that would imply severe harm to others: “Hubris disregards those who would have to live with the potential failure of the technologies in question” (Meyer and Uhle 2015: 5). Hubristic innovators neglect the dangers they impose on others by misjudging their abilities to reshape the world in the way they want.

Though our purpose is not to assess this objection in detail, it is important to note that its classical version runs into the same problem as the objection of the artificialization of nature. It relies on a too-strict separation, this time between humans and divine activity. In addition, the idea of “divine power” is based on a problematic theological assumption. For these reasons, a more nuanced version of the accusation of hubris relying on the idea of “natural forces” is more promising:

- (1) There are natural forces at work in the world.
- (2) Humans do not totally understand these forces, let alone control them.
- (3) If humans try to master these forces, unintended and harmful consequences will likely result.

The hubris objection is important in many subfields of IE. For our present argument, it illustrates the relevance of ethical theory. The accusation of hubris can be framed in different ethical perspectives. In brief, it could be framed as a consequentialist maxim to avoid dangers, which are beyond what could ever be gained in return. The main difficulty is to make sense of this approach in the field of innovation, which is characterized by high degrees of uncertainty regarding the consequences of one’s choices. It could also be framed as a deontological rule, based upon the protection of rights, which puts a clear limit on what can legitimately be changed. These rights could be human rights but could also be extended to future generations or to non-human beings with interests which are deemed as valuable (e.g. Heavey 2013).

Finally, it could be framed along a virtue ethics approach. It is interesting to note that virtue ethics has been an important way to frame arguments in subfields linked to climate ethics (see Gardiner 2011; Jamieson 2014; Vallor 2016). Humble innovators would be the most appropriate agents because they are the most likely to have the required courage of living with the total responsibility that comes with intentionally manipulating the climate system. However, those who display character traits such as humility and courage are unlikely to see such projects as appropriate and fitting. They accept a certain lack of control over the conditions of life and are aware of the limits of human actions. They act scrupulously, and with prudence and ease. They are aware of their fallibility (Hartman 2017).

The shift towards virtue ethics might well extend to other subfields of IE. In facing the requirement to imagine a new concept of responsibility, some theorists come to address the character traits of innovators. On the one hand, the focus on the virtues and vices could be an instrument to address the ethical motivation to act (working as an instrument towards the realization of ethical goals otherwise defined and justified). On the other hand, it could indeed represent an ethical theory to justify these goals in a more fundamental way (Hursthouse and Pettigrove 2016). In both cases, virtues and vices of innovators are used as ways to operationalize moral responsibility.

The main point for the present argument is to identify the relevance of the underlying debate on the kind of normative ethics that is used. Depending on how we frame and justify the accusation of hubris, its normative implications will be substantially different. In subfields of IE related to environmental philosophy, the ethical grammar drawn upon to frame this kind of argument is telling of the potentiality of a virtue ethics approach. Because traditional understandings and operationalizations of responsibility do not seem to fit the task, there seems to be a shift towards a different approach seen as capable of fulfilling the task of justifying cautiousness when addressing innovation.

4.4 Implications for Responsible Innovation

This IE approach to the notion of responsibility brings four distinct lessons relevant for the literature on RI. Firstly, ethical reflections on new forms of responsibility, such as aggregative responsibility, inclusive responsibility and intergenerational responsibility (e.g. in form of total responsibility), enrich the understanding of the concept of responsibility on which RI relies and stresses its normative implications. Because of its historical records and its normative ambition, RI is a practice-oriented field of scholarship. Acknowledging and embracing the broader normative context might help to reduce the normative gap in the literature on RI (as proposed by van de Poel and Sand 2018).

Secondly, our approach might be useful in enriching the focus of debate. In the literature on RI, the question of responsibility is most importantly raised for actors within innovation processes. This focus allows the development of a practice-oriented approach to responsibility regarding innovation. The thematic approach developed here allows embracing a broader concept of responsibility. The example of “total responsibility” shows that this idea is not only process- or product-based. It might be operationalized in the context of R&D, but it is much broader in scope. It is broader in its capacity to integrate aggregative responsibility (e.g. on the tragedy of the commons,

Kahn 2014) and in sustaining a societal, regional or even global debate about individual and collective responsibility. To combine two dimensions of responsibility (RI-responsibility and IE-responsibility) makes explicit the crosslinks between the different fields. As exemplified with “total responsibility”, a general view of responsibility might represent the general normative framework in which a specific RI-responsibility should be developed.

Thirdly, this synergy might represent important resources when it comes to identifying mechanisms operationalizing responsibility or alternatives thereof (de Hoop et al. 2016). Following van de Poel and Sand, it might be argued that ascribing responsibility to innovators might only be possible in specific conditions. They argue that two types of responsibility are important: accountability and what they call “responsibility-as-virtue”. The general normative framework represented by IE-responsibility (such as “total responsibility”) could then be used as a set of resources to formulate principles fulfilling parts of the normative mission given to responsibility in RI. Furthermore, these resources might be drawn upon in designing institutional mechanisms fulfilling similar functions (such as an insurance scheme to address liability issues, as mentioned by van de Poel and Sand, § 3.3).

A fourth input of this investigation on the notion of responsibility is that it helps make the issue on the assumed ethical theory explicit. By revealing the possible uses of different normative ethics, it highlights that alternatives to the traditional consequentialist agent-centred approach of responsibility that dominates in the RI literature are possible. In this context, it might be used to explicitly address and make sense of the relevance of virtue ethics within RI (on this topic, see also Sand 2018). The first part is heuristic in nature. The main idea is to take note of these distinct underlying theories and to identify their strengths and weaknesses. The second part is more practical. Within RI, different theories are not necessarily heterogeneous theories in competition with each other; it is possible to conceive a pluralistic ethical approach that combines the virtues of the different ethical perspectives.

5 Conclusion

This paper makes the claim that the field of IE (as constructed here) might contribute to the development of the field of RI, with a particular focus on ethical issues raised by ongoing technological innovations.

Drawing on a first transversal issue within the field of IE (the artificialization of the world), we have highlighted that the discourse on RI would benefit from integrating the normative relevance of shifting conceptual boundaries between humans and machines, matter and mind, science and technology, and natural and artificial. In conceptualizing innovation, RI scholars should try to grasp the implications of some technological innovations on the very concepts we use to try to make sense of the world (“normal”, “natural”, “human” and so on). They should also recognize that, in turn, such concepts are not used in a morally neutral way, that is, that they have strong normative components. As explained when dealing with practical implications such as on the burden of proof, the use of these conceptual categories raises crucial ethical questions. IE scholars can contribute to enriching this discourse on RI by bringing their expertise on specific conceptual boundaries.

Drawing on a second transversal issue (the notion of responsibility), we have explained how IE could contribute to the discourse on RI by deepening the notion of responsibility with new forms of responsibility (such as aggregative responsibility, inclusive responsibility, and total responsibility). Our thematic approach proposes a broader concept of responsibility than the one which dominates in the literature on RI, by showing that some forms of responsibility, such as total responsibility, are not only process- or product-based. Our approach highlights here again, in line with our main objective, that ethical considerations need to be identified and developed in the discourse on RI, in this case by making explicit the normative ethics on which authors tend to rely. Although the RI literature is dominated by a consequentialist agent-centred approach, the example of the normative implications of total responsibility shows that a virtue ethics approach focusing on character traits, such as hubris and humility, could be highly relevant to analyzing current technological innovations.

Although we have mainly focused here on technological innovations in the context of the Synthetic Age, such as solar climate engineering projects, this approach could be broadened to other technological trends to make other transversal normative issues relevant for RI appear, such as digitalization, automation and datafication. Likewise, it could be possible to enrich the discourse on RI by going beyond the techno-economic paradigm and highlight normative issues raised by political, economic and social innovations. This sketches out an agenda for future research to strengthen the dialogue which we have started here between IE and RI.

Funding Information Open access funding provided by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Blok, V. (2018). Philosophy of innovation: a research agenda. *Philosophy of Management*, 17(1), 1–5.
- Blok, V. (Forthcoming). “Towards an ontology of innovation: on the new, the political-economic dimension and the intrinsic risks involved in innovation processes”. In Michelfelder, D., & Doorn, N. *Routledge handbook of philosophy of engineering*. London/New York: Routledge.
- Blok, V., & Lemmens, P. (2015). The emerging concept of responsible innovation. Three reasons why it is questionable and calls for a radical transformation of the concept of innovation. In B.-J. Koops et al. (Eds.), *Responsible innovation 2. Concepts, approaches, and applications* (pp. 19–35). Dordrecht: Springer.
- Bontems, V. K. (2014). What does Innovation stand for? Review of a watchword in research policies. *Journal of Innovation Economics*, 15(3), 39.
- Bloodworth, M. N. (2017). Sport, society, and anti-doping policy: an ethical overview. *Medicine and Sport Science*, 62, 177–185.

- Boucher, Olivier et al. (eds), "Clouds and aerosols". 2013. In *The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change* (eds. Thomas F. Stocker et al.). Cambridge: Cambridge University Press, 571–657.
- Bourg, D., & Whiteside, K. (2010). *Vers une démocratie écologique. Le citoyen, le savant et le politique*. Paris: Seuil.
- Bourg, D., et al. (2017). *Inventer la démocratie du XXI^e siècle. L'Assemblée citoyenne du futur*. Paris: Les Liens qui Libèrent.
- Brundage, M. (2014). Limitations and risks of machine ethics. *Journal of Experimental & Theoretical Artificial Intelligence*, 26(3), 355–372.
- Cohen, S. (2014). The ethics of de-extinction. *Nanoethics*, 8, 165–178.
- Crutzen, P. J. (2006). Albedo enhancement by stratospheric sulfur injections: a contribution to resolve a policy dilemma? *Climatic Change*, 77(3–4), 211–219.
- Dobson, A. (2003). *Citizenship and the environment*. Oxford: Oxford University Press.
- European Commission. 2014. Rome declaration on responsible research and innovation in Europe. https://ec.europa.eu/research/swafs/pdf/rome_declaration_RRI_final_21_November.pdf. Accessed 1st June 2019.
- Fabian, C., & Fabricant, R. (2014). The ethics of innovation. *Stanford Social Innovation Review*. https://ssir.org/articles/entry/the_ethics_of_innovation.
- Fabris, A. (2018). *Ethics of information and communication technologies*. Dordrecht: Springer.
- Frisch, D. (1993). Reasons for framing effects. *Organizational Behavior and Human Decision Processes*, 54, 399–429.
- Frogneux, N. (2015). Responsabilité (philosophie). In D. Bourg & A. Papaux (Eds.), *Dictionnaire de la pensée écologique* (pp. 872–874). Paris: PUF.
- Gardiner, S. M. (2011). *A perfect moral storm. The ethical tragedy of climate change*. Oxford: Oxford University Press.
- Giubilini, A., & Sanyal, S. (2015). The ethics of human enhancement. *Philosophy Compass*, 10(4), 233–243.
- Godin, B. (2009) "Innovation: the history of a category". *Project on the Intellectual History of Innovation*, Working Paper No. 1. Montréal: INRS. <http://www.csiic.ca/PDF/IntellectualNo1.pdf>.
- Godin, B. (2015). *Innovation Contested. The Idea of Innovation over the Centuries*. New York: Routledge.
- Gordijn, B., & Cutter, A. M. (Eds.). (2014). *In pursuit of nanoethics*. Dordrecht: Springer.
- Gremmen, B., Blok, V., & Bovenkerk, B. (2019). Responsible innovation for life: five challenges agriculture offers for responsible innovation in agriculture and food, and the necessity of an ethics of innovation. *Journal of Agricultural and Environmental Ethics*, 32(5–6), 673–679.
- Hartman, L. M. (2017). Climate engineering and the playing god critique. *Ethics & International Affairs*, 31(3), 313–333.
- Häyry, H. (2018). Ethics and cloning. *British Medical Bulletin*, 128(1), 15–21.
- Heavey, P. (2013). Synthetic biology ethics: a deontological assessment. *Bioethics*, 27, 442–452.
- Holy-Luczaj, M., & Blok, V. (2018). How to deal with hybrids in the anthropocene? Towards a philosophy of technology and environmental philosophy 2.0. *Environmental Values*, 28(3), 325–345.
- de Hoop, E., Pols, A., & Romijn, H. (2016). Limits to responsible innovation. *Journal of Responsible Innovation*, 3(2), 110–134.
- Hursthouse, R., & Pettigrove, G. (2016). "Virtue ethics". In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition).
- Jamieson, D. (2010). Climate change, responsibility, and justice. *Science and Engineering Ethics*, 16(3), 431–445.
- Jamieson, D. (2014). *Reason in a dark time: why the struggle against climate change failed – and what it means for our future*. Oxford/New York: Oxford University Press.
- Jamieson, D., & Di Paola, M. (2016). Political theory for the anthropocene. In D. Held & P. Maffettone (Eds.), *Global political theory* (pp. 254–280). Cambridge: Polity.
- Kahn, E. (2014). The tragedy of the commons as an essentially aggregative harm. *Journal of Applied Philosophy*, 31, 223–236.
- Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press.
- Katz, E. (2015). Geoengineering, restoration, and the construction of nature: oobleck and the meaning of solar radiation management. *Environmental Ethics*, 37(4), 485–498.
- Lee, K. (1999). *The natural and the artefactual: the implications of deep science and deep technology for environmental philosophy*. New York: Lexington Books.
- Lichtenberg, J. (2010). Negative duties, positive duties, and the 'new harms'. *Ethics*, 120(3), 557–578.
- Lin, P., Abney, K., & Bekey, G. A. (Eds.). (2014). *Robot ethics: the ethical and social implications of robotics*. Cambridge: MIT Press.

- McKibben, B. (1989). *The end of nature*. New York: Random House.
- Meyer, K., & Uhle, C. (2015). "Geoengineering and the Accusation of Hubris" *THESys Discussion Paper 3* (pp. 1–15). Berlin: Humboldt-Universität zu Berlin.
- Minerva, F. (2018). The ethics of cryonics. In *Is it immoral to be immortal?* New York: Palgrave Macmillan.
- Mittelstadt, B., & Fliridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, 22(2), 303–341.
- de Morais, J. C. B., & Stückelberger, C. (Eds.). (2014). *Innovation ethics. African and global perspectives*. Geneva: Globethics.net and African Innovation Foundation.
- Owen, R., et al. (2013). A framework for responsible innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible innovation* (pp. 27–50). Chichester: Wiley.
- Pavie, X. (2018). *L'innovation à l'épreuve de la philosophie*. Paris: PUF.
- Pellé, S., & Reber, B. (2015). Responsible innovation in the light of moral responsibility. *Journal on Chain and Network Science*, 15(2), 107–117.
- van de Poel, I., & Sand, M. (2018). Varieties of responsibility: two problems of responsible innovation. *Synthese*, 1–19. <https://doi.org/10.1007/s11229-018-01951-7>.
- Preston, C. J. (2012). Beyond the end of nature: SRM and two tales of artifice for the anthropocene. *Ethics, Policy & Environment*, 15(2), 188–201.
- Preston, C. J. (Ed.). (2017). *Climate justice and geoengineering: ethics and policy in the atmospheric anthropocene*. London/New York: Rowman & Littlefield.
- Preston, C. J. (2018). *The synthetic age: outdesigning evolution, resurrecting species, and reengineering our world*. Cambridge/London: The MIT Press.
- Risse, M. (2017). Responsibility and global justice. *Ratio Juris*, 30(1), 41–58.
- Robock, A. (2016). Albedo enhancement by stratospheric sulfur injections: more research needed. *Earth's Future*, 4(12), 644–648.
- Rolston III, H. (2017). The anthropocene! Beyond the natural? In S. M. Gardiner & A. Thompson (Eds.), *The Oxford handbook of environmental ethics* (pp. 62–73). Oxford: Oxford University Press.
- Sand, M. (2018). The virtues and vices of innovators. *Philosophy of Management*, 17(1), 79–95.
- Shah, M., Bhat, M., & Davin, J. (2014). *Nanotechnology applications for improvements in energy efficiency and environmental management*. Hershey: IGI Global.
- Sovacool, B. K., & Dworkin, M. H. (2014). *Global energy justice*. Cambridge: Cambridge University Press.
- Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580.
- Strawser, B. J. (Ed.). (2013). *Killing by remote control. The ethics of an unmanned military*. Oxford: Oxford University Press.
- Tate, M., Johnstone, D., & Fielt, E. (2017). "Ethical issues around crowdwork: How can blockchain technology help?". In *Proceedings of the 28th Australasian Conference on Information Systems (ACIS 2017)*, 4–6 December 2017, Hobart, Tasmania. <https://eprints.qut.edu.au/115042/>.
- Timmermans, J., & Blok, V. (2018). "A critical hermeneutic reflection on the paradigm-level assumptions underlying responsible innovation". *Synthese*: 1–32. <https://doi.org/10.1007/s11229-018-1839-z>.
- Trebilcock, M. (2015). *Dealing with losers: the political economy of policy transition*. Oxford: Oxford University Press.
- Vallor, S. (2016). *Technology and the virtues: a philosophical guide to a future worth wanting*. Oxford: Oxford University Press.
- Van den Hoven, J. (2013). Value sensitive design and responsible Innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible innovation* (pp. 75–84). Chichester: Wiley.
- Von Schomberg, R. (2012). Prospects for technology assessment in a framework of responsible research and innovation. In M. Dusseldorp & R. Beecroft (Eds.), *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden* (pp. 39–61). Wiesbaden: Springer.
- Von Schomberg, R. (2013). A vision of responsible research and innovation. In R. Owen, M. Heintz, & J. Bessant (Eds.), *Responsible innovation* (pp. 51–74). Chichester: Wiley.
- Von Schomberg, L. and Blok, V. 2019. "Prospects for technology assessment in a framework of responsible research and innovation" *Philosophy & Technology*: 1–15 (early view).
- Woolgar, S. (1985). Why not a sociology of machines? The case of sociology and artificial intelligence. *Sociology*, 19(4), 557–572.
- Ziegler, R. (2015). Justice and innovation – towards principles for creating a fair space for Innovation. *Journal of Responsible Innovation*, 2(2), 184–200.