TUM School of Social Sciences and Technology
Technische Universität München

THE POLITICS OF INNOVATION:

ESSAYS ON RESPONSIBLE INNOVATION AND THE ROLE OF PARTICIPATION IN ECONOMICS, BUSINESS, AND PUBLIC POLICY

Johann Jakob Häußermann

Vollständiger Abdruck der von der TUM School of Social Sciences and Technology der Technischen Universität München zur Erlangung eines Doktors der Philosophie (Dr. phil.) genehmigten Dissertation.

Vorsitzender:

Prof. Dr. Dirk Heckmann

Prüfende der Dissertation:

Prof. Dr. Christoph Lütge

Prof. Dr. Martina Schraudner

Die Dissertation wurde am 27.09.2021 bei der Technischen Universität München eingereicht und durch die TUM School of Social Sciences and Technology am 23.03.2022 angenommen.

ABSTRACT

Given the exceptional, positive role of innovation for economic and social progress, the underlying normative complexity has received only limited attention so far. Rather, a favourable view that considers innovation, mostly implicitly, as something good per se dominates the general understanding both in business and politics as well as in innovation research. In three independent publications, this dissertation explores the opportunities of responsible innovation and enables a more differentiated appraisal of the normative dimensions of innovation. In the face of increasingly complex ramifications of innovation and intensifying societal tensions, the dissertation contributes to the urgent re-politicisation of the concept.

ZUSAMMENFASSUNG

Angesichts der besonderen positiven Bedeutung von Innovation für gesellschaftlichen Fortschritt wurde der normativen Komplexität des Konzepts nur wenig Aufmerksamkeit geschenkt. Vielmehr dominiert ein wohlwollender Blick, der Innovation – meist implizit – als etwas Gutes an sich betrachtet, das allgemeine Verständnis sowohl in Wirtschaft und Politik als auch in der Innovationsforschung. In drei eigenständigen Artikel untersucht diese Dissertation die Möglichkeiten verantwortlicher Innovation und ermöglicht damit eine differenziertere Bewertung der normativen Dimensionen von Innovation. Angesichts zunehmend komplexeren Auswirkungen von Innovationen und sich zuspitzenden gesellschaftlichen Verwerfungen kann die Dissertation so einen Beitrag zur dringenden Re-Politisierung des Konzepts leisten.

CONTENTS

1. Introduction

Over the past centuries, technological change and new organisational and social practices have yielded vital advancements for both people and societies and continue to be drivers of economic and social progress today. Or, in the words of B. Godin (2008): Innovation is the emblem of modern societies. Yet, over the last decades it has become more and more evident that a solely favourable view of innovation as an "instrument for growth" provides an incomplete picture. Not only have so-called unintended negative consequences of new technologies increasingly entailed in part dramatic harms (Biggi and Giuliani 2021) and social and ethical impacts grown increasingly salient in light of the deep societal embedding of new technologies (Philbeck, Davis, and Larsen 2018), but also from an economic perspective the mono-causal relationship between innovation and economic growth is subject to greater questioning (Soete 2013), and challenges from climate change require a fundamental reorientation towards planetary boundaries (Rockström et al. 2009) as the basis for human life and thus for innovation. This highlights, amongst other things, the normative complexity of innovation, which has been neglected, if not ignored, in both academic debate as well as in practice in research and development, business, and politics – given the wealth effects of technological change. In other words, to this day a so-called "pro-innovation bias" (Rogers 1962) dominates the general understanding of innovation, impeding a more balanced normative understanding.

This dissertation seeks to draw a normatively differentiated picture of innovation against the background of the concept of responsible innovation (Burget, Bardone, and Pedaste 2017). To the extent that responsible innovation aims to align technological change with societal needs, perspectives, and values, and represents the most common approach today to systematically open up a predominantly technical-economic conceptualisation of innovation to societal aspects, it provides an adequate framework for illuminating the different normative dimensions of innovation. To derive a normatively more nuanced picture, three points of intersection with neighbouring disciplines and research fields which have been neglected in the debate on responsible innovation so far are being systematically explored. The three articles forming this dissertation firstly examine the normative complexity from the perspective of economic theory and derive a corresponding approach for responsible innovation, secondly illuminate the role of companies for responsible innovation and develop a corresponding governance framework from a business ethics perspective, and thirdly address the relationship between acceptance research and responsible innovation by

using empirical data. This way, the prospects of responsible innovation and its central governance mechanism based on participation shall be examined and enhanced by arguments and insights from three related research fields. Or, put differently, the dissertation seeks to demonstrate the demand as well as the prospects for responsible innovation using the terms and concepts of economic theory, business ethics, and acceptance research. Otherwise, responsible innovation may run the risk of eventually evolving into a topic in its own right – but without significant and effective reference to theories, concepts, and practices critical to innovation. This dissertation thus aims not only to contribute to the further development of responsible innovation, but also to ensure that the term "responsible innovation" no longer appears to be an oxymoron in the future (Delvenne 2017). Instead, responsible innovation should turn into a tautology: Innovation is an inherently social project about how society wants to progress through new scientific and technological means. The negotiation of costs and benefits, opportunities and risks, and their distribution is, simply put, nothing else than the question of what kind a society wants to be. By re-politicising the concept of innovation, this dissertation intends to help reconcile two major current concerns: innovation on the one hand, and social cohesion on the other. In view of the urgent challenges posed by growing societal polarisation and fragmentation and even crises of democracy (Przeworski 2019), it seems all the more important to not only focus on the influence of technological change thereon, but also to re-evaluate its functional role for societal stability and cohesion as a whole (Mokyr 1990). To this end, this dissertation contributes to transforming responsible innovation (back) from an oxymoron into a tautology.

The dissertation is structured as follows: Following this introduction, the second chapter first provides a general overview of the context of innovation [2.1] by highlighting the immense positive role of innovation today [2.1.1], presenting a brief history of the concept of innovation [2.1.2], and giving an introduction to the current field of innovation studies [2.1.3]. The second part of the chapter begins by discussing the dark side of innovation and its consideration in current research [2.2.1], before a general overview of the normative dimensions of innovation is introduced [2.2.2]. To the knowledge of the author this includes the first attempt of a general systematisation of a new field of innovation ethics. The third chapter provides an introduction of the concept of responsible innovation [3.1] and introduces the research questions of the three individual articles against the background of the current state of research in the context of responsible innovation [3.2]. Against this background, the following chapter entails the three individual articles of this dissertation [4.1-4.3]. Given the nature of its

subject a highly interdisciplinary research approach is necessary which combines perspectives from philosophy, ethics, social sciences, and economics as well as different theoretical, conceptual, and empirical methods. The three articles constituting the dissertation are in this sense a theoretical [4.1], a conceptual [4.2], and an empirical case study [4.3]. Building on this, chapter 5 briefly summarizes the most important results [5.1], discusses them with respect to the general implications for responsible innovation [5.2], and points to implications for both further research [5.4] as well as practice [5.5]. The dissertation ends with a reflection of its general findings in the light of current societal developments and debates.

2. Innovation is good, isn't it? On the dark side of innovation

This chapter provides an introduction into the dissertation's subject by explaining the general context as well as developing a first systematization of a new field of innovation ethics. The chapter consists of two parts, the first of which provides a more general introduction to innovation [2.1], while the second one focuses more closely on its "dark side" and its normative dimensions [2.2]. The first part is structured as follows: After a brief section highlighting the key role of innovation for modern societies and social progress [2.1.1], I will give a brief historical introduction into the concept of innovation [2.1.2] and present key approaches to researching innovation by recapitulating the current state of innovation studies [2.1.3]. The second part first summarizes the current research approaches investigating normative aspects of innovation [2.2.1], before developing a novel systematization based on four dimensions of a circle of innovation ethics [2.2.2].

2.1 A brief introduction to innovation

2.1.1 The positive impact of innovation

Science, technology, and innovation are a fundamental part of modern life. Part of the reason for this is the crucial role that new knowledge and technological developments have played in the development of modern societies and the significant contribution they have made to continuous progress in recent centuries. Although this has primarily led to growing economic prosperity in Western countries since the beginning of the 20th century and especially in its second half, one can also observe from a global perspective not only a clear upswing in so-called developing countries, but also a drastic improvement in living conditions overall. Whereas in 1990, 1.9 billion people, or 36% of the world's population, lived in extreme poverty, only 730 million people, or 9.9% of the world's population did so in 2015. This trend and its historical background become even stronger when we take yet another look at the period between 1820 and 2015, which highlights in particular the major positive role of the second half of the 20th century in the decline of global poverty (fig. 1).

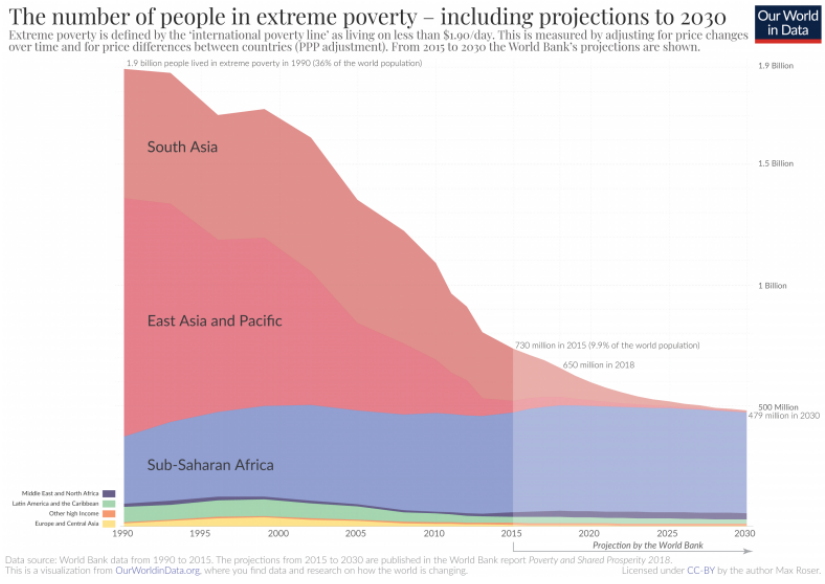

Fig. 1: Number of people in extreme poverty from 1990 to 2015 (Source: *Our World in Data*).

Assuming that a look at health improvements, e.g. in terms of life expectancy or infant mortality rates, allows for a more direct inference of the impact of innovation, these may reinforce the observation of the overall progress, especially in the second half of the 20th century: Global life expectancy increased from just over 30 years in 1900 to

over 70 years in 2019 (fig. 2), while infant mortality worldwide decreased from over 20% in the mid-20th century to under 5% in 2019 (fig. 3).

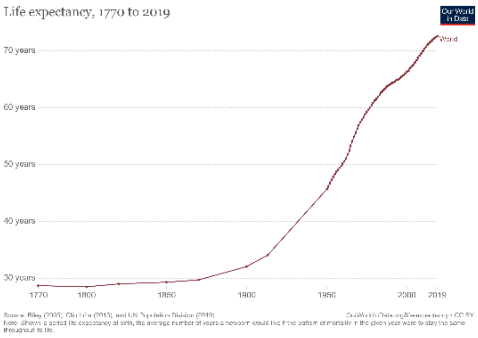

Fig. 2: Global life expectancy 1770 to 2019 (Source: *Our World in Data*).

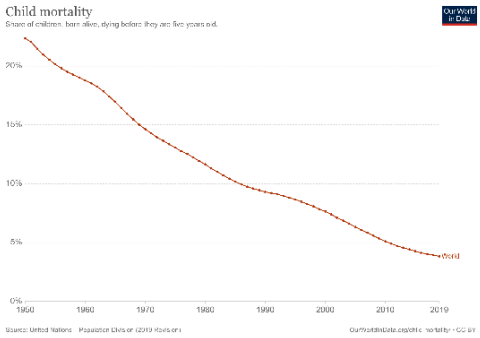

Fig. 3: Global child mortality 1950 - 2019 (Source: *Our World in Data*).

At the same time, not only has the gross domestic product per capita increased drastically worldwide since 1820 and again especially since 1950 (fig. 4), but also the number of registered patents as a very rough indicator of innovation (illustrated here by the example of the United States of America) grew continuously with the beginning of the 20th century (fig. 5).

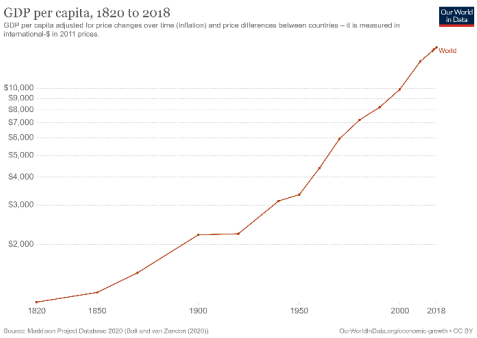

Fig. 4: Global GDP per capita 1820 - 2018 (Source: *Our World in Data*).

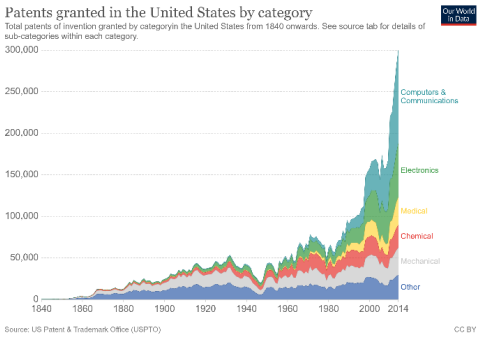

Fig. 5: Patents in USA 1840 - 2014 (Source: *Our World in Data*).

It is important to note at this point that the rather approximate description of the developments ignores the at times drastic discrepancies between different regions of the world as well as between individual countries. These are nevertheless of utmost importance for an in-depth analysis of these developments. However, what the highly

simplified juxtaposition is meant to reveal is the general progress and the positive effects on general living conditions worldwide, which can at least partly be explained by decisive scientific findings and technological developments of the past decades. The correlation of technological change and positive economic development is therefore a central element of a description of innovation and its importance for contemporary societies (Mokyr 1990; Grossmann and Helpmann 1991; Hasan and Tucci 2010). Despite the fact that the relationship between innovation and economic growth is complex (Verspagen 2006), it can be concluded from a retrospective perspective that innovation has been important for prosperity and overall welfare (Mokyr 1990). Moreover, innovation and technological change not only increased welfare but also helped advancing knowledge, changed human practices, and facilitated social coexistence. In short, innovations relieved human life and substantially enriched cultural life.

It appears almost trivial to use examples of individual innovations that have had a profound impact not only on economic, but also on social progress. Of course, innovations such as the printing press at the end of the 15th century or the industrial revolution in the 18th century, which was largely driven by the development of the steam engine, are truly emblematic. But also in more recent times, innovations such as genome editing based on the so-called CRISPR/Cas9 method or the technological change triggered by digitisation can be attributed revolutionary significance. Here the example of so-called artificial intelligence (AI) may be particularly relevant. AI systems are already today widely used in almost all sectors of society, from manufacturing, agriculture, trade, finance, and medicine to government and public administration. Applications range from digital assistants such as chatbots, language translation tools, recommender systems of varying complexity in the consumer sector or professional contexts, to applications for autonomous driving or complex robotic systems and face recognition technologies.[1] Even if the wider effects of AI can only be assessed in the longer term, it can already be said today that AI-systems are not only profoundly changing existing business models, products, processes, procedures and practices across sectors, but are also having a profound economic and social impact. This includes not only platforms or information and communication services that enable new types of collaboration and allow to empower diverse individuals and groups, but

[1] This paragraph builds on the second article of this dissertation (Häußermann and Lütge 2021).

also, for example, the radical improvements that AI technologies enable in medicine or for climate and environmental protection.

In short, it seems fair to say that, historically, innovation has enabled economic prosperity as well as societal progress, especially for Western societies, but also globally. Innovations thus constitute a basic element of modern societies. Last but not least, this means that addressing current major challenges might not be possible without innovation (Witt 2016; J. Schot and Steinmueller 2018).

2.1.2 A brief history of the concept

But what does the concept of innovation, after all an emblem of modern societies, actually mean? Originating from ancient Greek καινοτομία, the word "innovation" etymologically stems from the Latin verb innovare meaning "to renew" but also "to come back to", primarily used in a legal context (Bontems 2014; Benoît Godin 2008; Benoit Godin 2015). Later, the term was revived particularly in the writings of Niccolò Machiavelli (1513) and Francis Bacon (1625), and thus enjoyed its first renaissance. From the very beginning, the concept had a political character, in that it aimed to change existing conditions, yet without initially making any reference to creativity or new knowledge (Benoît Godin 2008). According to its origins, innovation is thus caught between a revolutionary overthrow of the status quo on the one hand and a more conservative adherence to existing conditions on the other ("everything must change so that everything can stay the same"). In this respect, interestingly enough, the controversial character and the rejection of and resistance to innovation is part of the discourse on the concept very early on. With this in mind, it is no surprise that the sociologist Gabriel Tarde (1890, 1895, 1898, 1902) first made innovation a core concept for explaining social change in the late 19th century and presented the first theory of innovation. Innovation, according to him, is characterised primarily by imitation, which arises from the rejection of original inventions (Benoît Godin 2008). Thus, for Tarde, imitation is a key principle for society and its evolution. Central to Tarde's understanding of innovation to explain social change were ideas closely related to the concept of imitation, such as novelty, originality, imagination, and discovery - even though he never provided an explicit definition of these terms. According to Tarde, innovation happens mostly unconsciously through changes in practices and behaviours, although society as a whole tends to be conformist and conservative, adapting changes in terms of social cohesion through opposition and imitation (Bontems 2014). It was not until the middle of the 20th century that E. M. Rogers (1962) revived the idea of such a comprehensive meaning of the concept and developed another sociological theory of innovation as an explanatory model for social change. With the beginning of the 20th century, a sociologically-driven understanding of innovation was further established and a number of theories of innovation were developed and discussed (Ogburn 1922; Gilfillan 1935; Hart 1932). Core elements of these innovation theories are the overcoming of a dichotomy between imitation and innovation and the emergence of a linear and process-oriented understanding of innovation (Benoît Godin 2008). Accordingly, innovation is characterised by a linear sequence of invention and imitation, which eventually leads to innovation through the

successful introduction and use of an invention. Innovation in this sense is therefore an adopted invention. In this way, over the course of the 20th century, the development of a linear process model of innovation reinforces the role of novelty and thus also creativity, knowledge, and technology, all of which enable the initial invention at the beginning of the process. The increasingly important epistemic quality of innovation, in the sense of novelty and necessary uncertainty (Witt 1996), now plays a crucial role not least in dealing with unexpected, in the sense of unknowable, side-effects (Delvenne 2017). At the same time, innovation is described as a social and, to a certain extent, systemic process, insofar as there may be an individual inventor, but this inventor always draws on a social context and cultural heritage and, moreover, cannot possibly guarantee the successful adoption and societal use of his or her invention on his or her own. The growing attention that the concept receives in the course of the 20th century ultimately leads to an understanding of innovation that is still influential today and that prioritises the production of new knowledge and, in particular, the development of technological inventions alongside their successful use and adoption (Benoît Godin 2008). As a result, not only a linear and process-oriented conceptualisation of innovation comes to the fore, but also the dimensions of production of an invention on the one hand and its utilisation on the other are not contrasted, but inherently thought together within the framework of a linear understanding.

Alongside the sociologically and techno-sociologically driven development of innovation theories, economic perspectives on innovation evolved in the course of the 20th century and became increasingly influential for the general understanding of the concept. Although, unlike in sociology or psychology, change is not a central category for economic theories, which instead deal with equalities and equilibria (Veblen 1898), the profound transformations brought about by increasing mechanisation, the industrial revolution and the Great Depression of the 1920s and 1930s prompted fundamental discussions in economics as well (Pigou 1932; Hicks 1939). The growing importance of technology was taken into account by an enhanced assessment and investigation of the deployment of technologies in companies and the economy, which eventually led to the introduction of the so-called production function at the end of the 1920s, linking the quantity produced of a good (output) to quantities of inputs. This way, the production function expresses the technological change in the terms of economic theory and attempts to represent it in the corresponding models. Most influentially, Robert M. Solow finally captured this in his definition of the residual of the production function as technological change (Solow 1957). This, in turn, laid the

foundation for what is until today a significant body of research on innovation from an economic perspective, namely the correlation of research and development with productivity measures. It was through this conceptualisation of the impact of technological change on firm productivity that innovation increasingly found its way into mainstream economics in the course of the 20th century. At the same time, however, the path that innovation has taken in economics cannot be understood without the influence of J. A. Schumpeter and the theory of evolutionary economics. Famously, Schumpeter described capitalist economics as a mode of creative destruction and thus made innovation the core element of the functioning of capitalist economies (Schumpeter 1934, 1942, 1947). The constant renewal of existing structures and production methods, i.e., continuous innovation activities, not only introduce new products and practices, but also keep the economic system as a whole alive on the basis of precisely this mechanism. Accordingly, Schumpeter distinguishes five different categories of innovation based on a classification by D. Ricardo: 1) introduction of a new good; 2) introduction of a new method of production; 3) opening of a new market; 4) conquest of a new source of supply of raw materials or half-manufactured goods; and 5) implementation of a new form of organization. Although Schumpeter departs from sociological conceptualisations and differentiates between invention and innovation (Benoît Godin 2008), he similarly follows a linear and process-oriented understanding that emphasises the successful introduction and use of innovation alongside its production. Hence, even though he is often called the founder of innovation theory in the sense of technological innovation, he builds on existing sociological literature on innovation and, like other economists, conceptualises innovation essentially as a change in modes of production through new combinations of the available means of production. This paved the way, through the increasing influence of the economic perspective, for the conceptualisation of innovation as commercialised innovation, which is still authoritative today (Benoît Godin 2008). According to such a linear understanding of innovation, basic research and applied research as necessary steps in the development of new knowledge and new technologies are followed by the production and manufacture of innovations, which are then commercialised and introduced into society via markets and put to use there. This so-called "technology-push" model (in contrast to a "demand" or "market-pull" perspective that came into focus later) not only had a decisive influence on the sociological theories of innovation and economic research that have gained considerable momentum in the course of the 21st century, but also represented an important point of reference for innovation and technology policy since the end of the Second World War (Bush 1945). Against this background, i.e. parallel to the

increasing technologisation of production and the economy as a whole, innovation has been increasingly shaped by an economic perspective that sees innovation as an efficient instrument for increasing productivity and thus for economic progress and growth: „Many factors contributed to this shift: the political and economic context, the industrial and consumer revolutions, the impacts of technologies on individuals and societies, technology as a source of economic growth and productivity and, above all, the institutionalization of technological invention via patent laws, and industrial development through R&D laboratories. To many, innovation thus became an industrial and economic affair" (Benoît Godin 2008, 44). Until today, a predominantly economically driven view of the concept of innovation continues to dominate the scientific analysis, its practical application, as well as the political approach and the societal debate. This not only involves an understanding of innovation that is often technology-driven, but also focuses on the way in which new knowledge and new technologies are introduced and used in society through commercialisation by companies in international and globally organised markets.[2] Innovation is thus considered as prime strategy for gaining a competitive advantage and generating sustained profits (Witt 2016; Acs et al. 2012; Baumol 1990), while from a political perspective innovation serves as a solution to seemingly every problem (Bontems 2014) and is considered a permanent desideratum (Witt 2016). Or in the words of D.A. Schön, innovation came to be seen as the instrument of growth, and growth as the occasion for and object of innovation (Schön 1967, 54 following Godin 2008).

[2] This description of the predominant innovation paradigm is not meant to ignore developments that attempt to go beyond it, such as in the field of social, sustainable, green, or inclusive innovation. Such critical voices, which aim to question and expand the one-dimensional focus on technology and marketisation, should by no means be forgotten at this point, as they are in fact of particular importance for the topic of this dissertation. However, in so far as the mainstream approach and the generally prevailing understanding of the concept of innovation is being discussed here, a correspondingly more narrow presentation has been chosen.

2.1.3 Researching innovation today: The field of innovation studies

Against the background of this historical qualification of a conceptualisation of innovation, different strands of research can be distinguished today. In this context, the more recent academic debate on innovation is not only a multi- and interdisciplinary endeavour from the very beginning (Fagerberg and Verspagen 2009), substantially shaped by sociological and economic perspectives, but is also closely entangled with political missions and organisational formations (Fagerberg, Martin, and Andersen 2013; Fagerberg, Mowery, and Nelson 2006; Fagerberg and Verspagen 2009; B. R. Martin 2012). Building on the view of innovation as a technology-driven linear development process established in the course of the 20th and until the middle of the 21st century, systemically oriented perspectives on innovation came to the fore in the second half of the 21st century and especially from the 1960s to 1990s. While in a form of knowledge generation referred to as Mode-1 (Gibbons et al. 1994) new knowledge and innovations emerge along disciplinary boundaries and strict academic frameworks, so-called Mode-2 forms of knowledge production take account of a dynamically changing environment for science and technology. To this end, they reflect an increasing application orientation of knowledge production, which not only focuses knowledge and technology development more on specific contexts and challenges, but also contrasts a stronger "market-pull" dynamic with a "technology-push". In addition to an increasing application orientation, the emerging Mode 2 paradigm of knowledge production is characterised in particular by its stronger transdisciplinarity and "social distribution", whereby a so-called "co-production of knowledge" moves to the fore, which, beyond classic knowledge producers such as academia, involves various individuals and organisations and their heterogeneous knowledge resources in development, research and innovation processes (Nowotny, Scott, and Gibbons 2001, 2003; Gibbons 2000; Gibbons et al. 1994). This means that not only new challenges, questions, and equally new responses and solutions enter the horizon, but it also leads to a transformation of the previous linear organisation of innovation. This shift towards a more complex set of actors and processes required for the creation of innovations is reflected in a systemic view of innovation in the scientific debate. In particular, the highly influential theory of so-called national innovation systems (Lundvall 1992; Nelson 1993) has been instrumental in broadening the prevailing linear perspective on innovation by turning the focus to networks, interactions, and underlying systemic conditions. Until today, the model of national innovation systems constitutes an influential basis for understanding innovation against the background of the relationships and cooperations between the

diverse actors of an innovation system. Against this background, the so-called Triple Helix model (Etzkowitz and Leydesdorff 2000) was developed, which centres on the interactions between academia, business, and politics in order to analyse innovation. Building on this, further concepts were developed as part of the current debate on innovation from a systemic-driven perspective, each of which places a specific emphasis on dealing with the changing general parameters for innovation. Particularly influential in this regard has been the work of E. Carayannis and D. Campbell on a so-called Mode 3 paradigm of knowledge and innovation production (Carayannis and Campbell 2012b) together with a Quadruple Helix model (Carayannis and Campbell 2012a; Arnkil et al. 2010), which particularly stresses the role of civil society perspectives as a driver and lay knowledge as added value for new knowledge, services and products on the one hand, and to address growing legitimacy requirements and acceptance concerns, on the other. A step further goes the so-called Quintuple Helix model (Carayannis, Barth, and Campbell 2012; Carayannis and Rakhmatullin 2014), which, in view of the immense challenges of the global climate crisis, highlights the ecological in addition to the societal context. Approaches such as "user innovation" (Hippel 1988) or "open innovation" (Chesbrough, Vanhaverbeke, and West 2006) also play an important role in this context (Häußermann and Heidingsfelder 2017). Thereby, the sociologically driven and systemic approach forms a major research strand for the analysis of innovation today (Fagerberg, Mowery, and Nelson 2006). Finally, the concept of so-called innovation ecosystems is connected to this (Schroth and Häußermann 2018; Adner and Kapoor 2010; Gomes et al. 2018), which combines systemic approaches with perspectives of more traditional innovation management and their focus on value creation, among other things (Autio and Thomas 2014). This is achieved by breaking down and merging traditional boundaries between science and business ecosystems (Valkokari 2015).

This approach provides a link to other primary strands of innovation research, which analyse the general conditions, opportunities, and challenges for innovation from an economic or management perspective in light of economic, entrepreneurial, and organisational contexts. Economic research today is still largely influenced by a Schumpeterian conceptualisation of innovation, as it starts from the central role of companies as decisive actors in the creative process of developing and introducing innovations (Fagerberg and Verspagen 2009). In addition to the statistical measurement of innovation activities via various indicators such as patent filings or GDP growth, which continue to be a core element of innovation research and innovation policy today, especially against the background of the early entanglements of innovation research and innovation policy (Benoît Godin 2008; Fagerberg and

Verspagen 2009), economic analyses focus, for example, on the potential and impact of new products, services, and new production methods on markets and business models or on the organisational environment for innovativeness and the impact of new technologies on companies. Issues in economic innovation research therefore concern, on the one hand, organisational questions of innovation management about, for example, required knowledge, necessary entrepreneurial capacities, the involvement of external actors, the role of leadership, and organisational cultures or the design of innovation processes. At a different level, economic research strands explore the effects of innovation on particular markets and societies, e.g., with regard to employment effects in particular sectors, growth effects and forecasts, the distribution of costs and profits, or the representation of complex correlations via suitable (mathematical) models. Similar to the summary presented here, J. Fagerberg (2006) distinguishes four current research strands in innovation research:

1) Innovation in the making
2) The systemic nature of innovation
3) How innovation differs
4) Innovation and performance

In line with the preceding introduction of the sociological perspective on the systemic nature of innovation, Fagerberg summarises the part of innovation research that examines "the roles of institutions, organisations, and actors in this process at the national and regional level" under "the systemic nature of innovation" and the comparison of different systems with regard to temporal, spatial, political or industry-specific variations under "how innovation differs" (Fagerberg 2006, 4). Fagerberg describes the above-mentioned economic perspective on innovation in the sense of entrepreneurial questions of innovation management under the heading "innovation in making", while the economic analysis of the different effects is summarised under "innovation and performance".

In conclusion, it can thus be said that innovation research today is primarily shaped by sociological and economic research strands which, along specific historical developments of the concept of innovation, have departed from a linear notion in the sense of a "technology push" and, in the course of the second half of the 21st century, have shifted towards a systemic understanding based on the complex relationships between heterogeneous actors and dynamic processes and cooperation structures for the development and introduction of innovations, which tend to be driven by the societal demand side ("society/market-pull"). Furthermore, the brief description of the

historical development of the concept of innovation makes clear that the sociological perspective of science and technology studies on innovation today is also significantly shaped by an economic perspective (J. C. Schmidt 2007). This finds expression not only in the (implicitly) underlying concept of innovation in the sense of "commercialised innovation", but also in the often unquestioned normative evaluation of innovation in general. It is not least due to the economic character of innovation research and the economically driven view of innovation as a whole that leads to innovation being regarded as something positive, valuable, to be generally endorsed and, all in all, as something "good". Apart from a few exceptions in innovation research (Georgescu-Roegen 1976; Nelson 2011, 1977; Justman and Teubal 1986), the general "pro-innovation bias" (Rogers 1962) appears historically reasonable and to some extent justified in view of the enormous positive effects on the economy and the economy (Mokyr 1990). However, a solely positive view of innovation is problematic not only in that it loses sight of the complexity of associated costs and benefits and possible negative effects as well as their distribution, but also operates at the theoretical-conceptual level with an under-complex and incomplete concept of innovation. In any case, the „pro-innovation bias“ (Rogers 1962), also referred to as „cult of innovation“ (Winner 2018), „gospel of innovation“ (Komlos 2016), „ceremonnial value of innovation“ (Schubert 2015a), „innovation as ideology“ (Benoît Godin and Vinck 2017) or „innovation fetish“ (David 2012), reveals not only a general underlying technoscientific optimism(J. C. Schmidt 2007) and a general equation of innovation with progress (Schubert 2015a), but the deficiency to adequately deal with the normative dimension of innovation. Even though the inherently destructive side of innovation was explicitly recognised by J. Schumpeter in his dictum of "creative destruction" (Schumpeter 1942) - and here even in the main noun for the central process of economic development - the failure to take this into account appears to go back to Schumpeter's historical observation and conclusion that innovation, in the long run and on average, improves the general welfare of the population (raising the “standard of life of the masses“ (Schumpeter 1942, 68). What remains unclear, however, is whether creative destruction can be considered legitimate at all: „Is what created in a process of change necessarily better than what is destroyed?“ (Schubert 2013; Mulgan 2016). With this in mind, it seems surprising to note that the destructive side of innovation has hardly ever received attention (Metcalfe 2001; Buenstorf et al. 2013a; Schubert 2013; Coad et al. 2021). This does not only refer to unintended consequences of new technologies or negative externalities in terms of environmental costs, but also to problematic effects from an economic perspective, e.g. with regard to diffusion and acceptance (Rogers 1962; Juma 2016; Tenner 1997) or market

developments that make innovation increasingly be viewed as not unconditionally positive from an economic perspective either (Soete 2013; Komlos 2016; Buenstorf et al. 2013b). Overall, it can be concluded that the pivotal role that innovation as technological change has played for economic prosperity has led to an increasing blindness to the normative intracies at different levels. Insofar as economic growth is generally regarded as a goal worth pursuing, and innovation seems to have proven its worth as a central instrument and primary strategy for achieving this, both at the level of society as a whole and at the corporate level, innovation has over time increasingly become a goal to be embraced per se, the normative dimensions of which do not seem to need to be questioned. Ignoring the normative complexity seems problematic, not least with regard to innovation and science and technology policy (Bontems 2014; Schubert 2015a; Soete 2019; Edler and Fagerberg 2017), which against this background also takes innovation as something inherently good, of which the more the better (Coad et al. 2021) and thus often focuses on the mere enabling and suitable framework conditions for innovativeness and innovation. Recent tendencies, however, show an increasing shift away from a "market-fixing" approach as normative justification for policy interventions towards demand-(Edler 2006), transformation- (J. Schot and Steinmueller 2018) and especially mission-oriented (Mazzucato 2016; European Commission 2018; Mazzucato 2017) approaches to innovation policy, which also address the question of the normative substance of value creation (Mazzucato and Perez 2014; Mazzucato 2018).

Against this background, this dissertation examines the "dark side" of innovation and addresses its normative complexity. The aim is not to contrast an (implicitly) one-sided positive assessment with a negative one, but rather to show approaches how a normatively more complete picture can be drawn from within the existing research strands. To this end, the following chapter attempts to provide an overview of the various normatively relevant dimensions of innovation. Though the interest and attention is growing, until now no attempt has been made to provide a comprehensive overview of the normative issues associated with innovation. While it is beyond the scope of this section of the dissertation to give a detailed explanation of the different levels and lay out a thoroughly exhaustive matrix, I will only be able to make a very first attempt to provide a preliminary systematisation of the different normative dimensions related to innovation.

2.2 The ethics of innovation: Mapping an emerging field of research

The previous chapter offered an overview of innovation as an object of research. I was able to show that across the various disciplinary and methodological approaches, a normative understanding is adopted, at least implicitly, that conceptualises innovation as something good per se. From a historical perspective, the so-called pro-innovation bias seems to be empirically plausible given the many improvements technological change brought about. This is especially true from an economic point of view, which puts emphasis on the positive effects of innovations for companies and national economies (Romer 1986; Aghion and Howitt 1992; Dosi et al. 1988; Verspagen 2006). But also from a broader societal perspective, a variety of different indicators can be used to demonstrate the tremendous positive effects, for example in terms of child mortality or global poverty (Coad et al. 2021). This ultimately led to innovation being viewed as a key driver and success factor for competitiveness at the level of firms and industries as well as at national levels (Lundvall 1992). However, the fact that the risks and potential negative effects of innovation have been pushed into the background in this light (Biggi and Giuliani 2021) and have hardly been taken into account in research (Dosi 2013) seems problematic. In short, an implicit pro-innovation bias (Rogers 1962; Sveiby, Gripenberg, and Segercrantz 2012) leads to a one-dimensional cult of innovation (Winner 2018) that equates innovation with progress (Schubert 2015a) and ignores the normative complexity of the concept of innovation and its key principle of creative destruction (Witt 1996).

2.2.1 On the dark side of innovation

In recent decades there have been several voices and signals that have clearly exposed the dark side of innovation (Jonas 1979; Beck 1986; Nelson 1977). Mostly, however, examples of this kind were understood as unique cases and exceptions that interpreted the potential risks of innovation from external causes rather than from the concept itself. In doing so, negative incidents are usually seen as so-called negative externalities of either individual technologies or economic growth as a whole, entailing risks and potential damage to e.g. human health or the natural environment. Well-known examples have almost become emblematic today: While asbestos was introduced in the 1950s as an innovative material that promised a remedy wherever protection from fire and collision was needed, by the late 1970s and 1980s its influence on the development of cancer among humans was discovered and all materials and products in which asbestos was now widely used were dismantled and destroyed (Tenner 1997; Witt 1996; Mokyr 2000; Benoît Godin and Vinck 2017). Or so-called chlorofluorocarbons (CFCs), a multi-purpose innovation in the chemical industry, which were used to replace potentially explosive chemicals in refrigerators or were used as lubricants, coolants, and fire extinguishing compounds, until it became apparent that the stability of the gas in the lower atmosphere contributed massively to the depletion of the ozone layer – which in turn led to an increase in ultraviolet radiation that subsequently raised the risk of skin cancer (Tenner 1997; Witt 1996; Juma 2016; Mokyr 1992). This problem resulted, as is well known, in the Montreal Protocol (United Nations 1989), a milestone in international environmental policy. Yet twelve of the most toxic chemicals ever invented, banned under the 2001 Stockholm Convention because of their harmful effects on human health and the environment, can still be found in the most remote areas of the world, such as deserts, the open ocean, and Arctic regions, where there is no human activity or chemical applications (Lohmann et al. 2007). Further examples of the list include for example the problematic increase of allergies such as asthma (which are somehow conspicuously absent in Amish communities (Holbreich et al. 2012)); the rapid increase of anaphylactic shocks (Lee et al. 2017); the prolonged decrease in human sperm counts (Levine et al. 2017); the potential discrimination caused by the increasing use of AI-based assistance systems in various fields of application (Whittaker, M., Crawford, K., Dobbe 2018) or the environmental costs of large language models natural language processing (NLP) (Bender et al. 2021); the rise of child-hood obesity (Ebbeling, Pawlak, and Ludwig 2002); the collapse of insect populations (Kunin 2019); pervasiveness of fluorinated chemicals in the environment

(Lim 2019), plastic waste accumulating in the oceans (Cressey 2016); the 'annihilation' of vertebrate populations in the context of the current human-induced mass extinction event (Ceballos, Ehrlich, and Dirzo 2017); the increasing share of Earth's surface made desolate and uninhabitable by human activity (e.g. Chernobyl, Fukushima, polluted battlefields such as Verdun in France (Bausinger, Bonnaire, and Preuß 2007; P. Cooper 2018), the Union Carbide disaster area in Bhopal with continuing groundwater contamination, nuclear waste disposal areas, former mining sites, or landmine fields); new pesticides such as DDT, new drugs such as thalidomide (Contergan) and rofecoxib (Vioxx), new techniques such as nuclear power generation and deep sea drilling, and so on (Coad et al. 2021; Binder and Witt 2019).[3]

Negative consequences are also becoming increasingly apparent at the economic level, for example with regard to the ecological effects of global capitalism (Giuliani 2018) or the increased incidence of cancers in most of advanced economies (Luzzati, Parenti, and Rughi 2018). But also in the field of finance, innovations have not only positive and growth-promoting effects in view of the calamitous contributions to the 2008 financial crisis (Dosi 2013): financial innovations such as sub-prime mortgages, collateralised debt obligations (CDO) and credit default swaps (CDS) all played a central role in creating the crisis, giving rise to a process of 'destructive creation' (Soete 2013; Biggi and Giuliani 2021). In another context, scholars have noted that, at times, technological change has also had detrimental effects on employment. Recently, for example, the effects of a broad deployment of artificial intelligence and other automation technologies have been discussed, having labour-replacing effects in a range of sectors and on different occupational profiles (Korinek and Stiglitz 2017; Autor 2015; Frey and Osborne 2017; Bughin et al. 2017; Bonin, Gregory, and Zieran 2015; Arntz, Gregory, and Zierahn 2017; Gregory, Salomons, and Zierahn 2019; Müller-Wieland et al. 2019). Other adverse economic impacts concern, for example, lock-in dynamics (Soete 2013), rent-seeking innovations through the deceiving of consumers (Akerlof and Shiller 2015) or monopolisation effects of data-driven business models and digital platforms (Kenney and Zysman 2016; Haucap 2018). Looking at the economic impact of ICT innovation, L. Soete (2013, 2019) argues that while it has made a large positive contribution to an overall increase in growth, it has increasingly turned innovation into a mechanism of "destructive creation", promoting unsustainable consumption growth. Accordingly, „the emergence of such service differentiation has also led to opportunities for cherry-picking: for selecting those

[3] Many examples of this section are based on Coad et al. (2021) and Biggi and Giuliani (2021).

segments of demand for profitable delivery at the expense of other less profitable segments, undermining as a consequence 'full' service delivery. As a result, many features of 'universal service' delivery associated with previous network service delivery have come under pressure. Their quality of delivery has become of lower quality, or in the worst case has even been discontinued" (Soete 2019, 138). However, by following the internal economic logic of destroying older product versions, ICT innovations have an unsustainable environmental impact and ecological footprint and thus promote an unsustainable form of consumption growth. Similarly, J. Komlos (2016) argues that innovation needs to be re-evaluated from an economic perspective due to the changes in the course of the 21st century: „Innovation's net value added to GDP or employment, –net of the negative externalities – has most likely diminished substantially in the twenty-first century. The destructive forces of creative destruction have gained the upper hand" (Komlos 2016).
In addition to examples of such technological and economic externalities, questions of diffusion, adoption, and acceptance of innovation also play a decisive role, insofar as these are not only an expression of normative entanglements, but controversies, resistance, and technological inertia also have social and political effects, for example on the stability of societies, and thus also fulfil a democratic function (Juma 2016; Tenner 1997; Mokyr 1994, 2000, 1992; Mokyr, Vickers, and Ziebarth 2015; Metcalfe 2001).

In view of these examples of possible negative effects, the fact that the dark side appears somewhat neglected in innovation research seems astonishing (Schubert 2013; Sveiby, Gripenberg, and Segercrantz 2012; Buenstorf et al. 2013b; Coad et al. 2021). For example, in a literature review of all articles in the EBSCO database with "innovation" in the title, Gripenberg et al. (2012) found that only 26 articles addressed possible unintended and undesirable consequences of innovation. This corresponds to a share of 1 per 1,000 and seems to have not changed since the 1960s (Gripenberg, Sveiby, and Segercrantz 2012). Although innovation, science and technology play a prominent role in the development of modern societies and negative externalities as well as ethical issues, especially in the context of specific technologies and their applications, have recently become the focus of increasing attention (Coad et al. 2021; Biggi and Giuliani 2021), so far no consolidated research area called innovation ethics exists. Instead, normative dimensions and ethical questions are addressed at different margins of the field of innovation research – at least in part, however, with growing attention, as will be shown in the example of the concept of responsible innovation in the next chapter. However, seldom the adverse impacts of innovation are seen in

context with and as inherent to the concept of innovation itself (Vincent Blok and Lemmens 2015; Witt 1996; Binder and Witt 2019). This appears particularly intriguing as "the dark side of innovation" has already been recognized early on by M. Schumpeter, one of the key figures in innovation research, coining the term of creative destruction (Schumpeter 1942; Schubert 2013).

2.2.2 Towards a new field of innovation ethics

2.2.2.1 Background

Before the next section presents an overview of the different normatively relevant dimensions of innovation against the backdrop of the current state of research, first the general approach to normativity in the context of innovation research will be described.

As already shown in the description of the development of today's social science-driven innovation research (see 2.1.3), constructivist concepts of the production of new knowledge and new technologies are mostly taken as a starting point. On the one hand, this is based primarily on sociological models of the so-called Mode 2-form of knowledge production, according to which a previously strictly hierarchical, disciplinarily organised, linear emergence of innovations oriented to classical actors in science is now determined by strongly contextualised and application-oriented processes of knowledge production, which in particular also involve societal needs and actors more directly in the development of innovations (Gibbons 2000; Gibbons et al. 1994; Funtowicz and Ravetz 1993). According to such a constructivist understanding of the emergence of new knowledge and new technologies, one can speak of "co-shaping", in the sense that societal perspectives and values are significantly involved in the shaping of new technologies and innovations from the very beginning (Bijker, Hughes, and Pinch 1987; Bijker and Law 1992; Yoshinaka, Clausen, and Hansen 2003; Ropohl 1979; Lenk and Ropohl 1987). In this sense of a social and ethical shaping of technology and innovation, normative dimensions are inherently inscribed in the processes of development and diffusion of new technologies and innovations. The task of normatively oriented science and technology studies is then, on the one hand, to identify, reflect on, and critically discuss implicitly embedded and hidden normative values, ideas, assumptions, or even biases. More practical approaches such as value sensitive design aim in this sense at a normatively reflected design of innovations oriented towards explicit values (Friedman and Henry 2019). Insofar as the normative analysis of innovation from a social science perspective in the course of the 1990s increasingly aimed at the early shaping of new technology in line with societal needs ("better technology in a better society") (Rip, Misa, and Schot 1995), the aim was also to minimise the risk of possible negative impacts at an early stage. However, to the extent that a constructivist understanding of a societal co-shaping of innovation was assumed, this also provided the basis for the establishment of a societal relationship of responsibility for potential adverse consequences (Jonas 1979; Lenk and Maring

1993). Last but not least, it is this notion that is essential for the further evolution of technology assessment into a concept of responsible innovation (Liebert and Schmidt 2015; Pellé and Reber 2015). In summary, it can thus be stated that current social science research on innovation does by no means describe innovation as ethically and normatively neutral (Brand 2020), but rather understands the normatively relevant societal ramifications of innovation against the backdrop of a constructivist understanding of systemically inscribed values and normative assumptions. This leads to the fact that concepts and approaches of technology assessment increasingly aim at making visible, reflecting, and finally actively shaping the normative dimensions of new technologies, insofar as their a-normative assessment is not possible: „It has become clear that a separation between scientific – and thus unbiased – analysis and subsequent normative analysis in the TA [technology assessment] process is untenable" (Bechmann et al. 2007, 13). This ultimately led to an opening of innovation research to diverse perspectives and a fundamentally participatory orientation (Jasanoff 2003), in particular including the concepts and methods of technology assessment (Abels and Bora 2004; Joss and Bellucci 2002; Hennen 1999, 2012) (see also 3.1.).

From an economic perspective, dealing with the normative dimension of innovation is not quite as explicit, at least the economic analysis of the normative levels of innovation has received little attention (Schubert 2015a) – although Schumpeter's (1942) influential conceptualisation of innovation already exposed a core normative dimension: the welfare gains of innovation are unevenly distributed of the economy and over time (Binder and Witt 2019). From standard neoclassical economic theory, the decisive normative criterion for evaluating innovations is the market test (Witt 2016; Mokyr 2000): If a corresponding demand exists and the innovation meets a certain need, e.g. by offering a new solution to a more or less societally relevant problem or by simply promising a better price or better quality, the innovation will be evaluated positively on the market and may not only be considered desirable but also as welfare improving. This kind of normative understanding is based on preference satisfaction theory of welfare of standard neoclassical economic theory, according to which the relevant normative criterion is provided by people's individuals preferences as revealed by their choices (Hausman, McPherson, and Satz 2017) (see also the first embedded paper of this dissertation). This forms the basis for societal evaluation on the basis of the so-called Pareto optimum, according to which that state of society is desirable in which it is not possible to improve the welfare (in terms of preference satisfaction) of one person without worsening the welfare of another. Although innovations and technological change as a whole quite obviously do not fulfil this

equilibirium criterion but rather lead to an unequal distribution of welfare gains and losses, they nevertheless seem desirable from an economic point of view. This is solely due to the compensation of the losers by the winners – even if this is only hypothetically possible in the sense of the Kaldor-Hicks criterion. Or in the words of Binder and Witt (2019, 352): „Under competitive conditions, the gains from an innovation are usually larger for the economy as a whole than the sum of the losses caused by pecuniary externalities. This is the very reason for why innovations do lead to a growth of per capita income (and a different way of saying that sometimes heavy damages to the welfare of some individuals do not exclude a positive average welfare effect of innovations). Because of this very fact it would, in principle, be possible to compensate all who lose out of the gains of the winners. Yet such compensations hardly ever take place. Nonetheless, when the mere possibility exists – [...] – this is often considered sufficient for claiming that society as a whole can realize a welfare gain through innovations." Not only does the condition of compensation seem to entail some problematic theoretical assumptions (such as, for example, the possibility of an interpersonal comparison of utility or the fact that preferences become endogenous to the innovations they are intended to evaluate (see also first paper of this dissertation)), but above all it just often seems not to occur and does not seem theoretically possible or meaningful. For although compensation in monetary terms may still seem conceivable, compensation in terms of utility, the actual criterion of the compensation principle, is much more difficult – and in most cases could also turn out considerably differently (Komlos 2016). In short, both the often invoked Kaldor-Hicks efficiency criterion and the Pareto criterion, both of which are supposed to provide normative justification from an economic perspective, do not seem to be applicable to the context of innovation (Komlos 2016; Schubert 2015a; Binder and Witt 2019; Schubert 2013; Mokyr 2000). As a consequence, it seems not only much more complicated than most economists would assume to demonstrate that innovation has an overall beneficial effect on society, but above all "the question of whether a possible increase in per capita output and living standards through continuing innovativeness is worth the risks involved" remains unsolved (Witt 1996, 114). Or put differently: "How serious must or can the privations be for how many, and for which, people, before the costs become too high for the innovations to be acceptable as a source of progress?" (Witt 1996, 116). Moreover, the current practice and implicitly guiding understanding of an economic-driven approach to innovation seems not only theoretically problematic, but also "hardly democratic insofar as the public is not asked in a referendum whether it wants the new technology or at what rate it would like to transition to the new technology. It is not asked whether we should

allow the destruction of the existing technologies and the concomitant misery of those whose labor is rendered superfluous by the innovation“ (Komlos 2016). While for Schumpeter it may have been a historically accurate observation that innovation in the long run and in the aggregate leads to an improvement of welfare, “[i]t is [...] now plain we have overestimated the productivity gains associated with technological change in the twentieth century. [...] This means the social costs of new techniques (as opposed to the costs captured in market prices) are systematically underestimated” (Mokyr 2014). In other words, although standard neoclassical economics based on welfare economics offers a central normative account for dealing with innovation – and this account is not only authoritative for the economic discussion of innovation, but also for policy-makers and their understanding of innovation as an "instrument for growth" – it proves to be at least problematic. At the very least, the mainstream economic perspective does not seem to adequately reflect the normative complexity of innovation theoretically and conceptually and thus does not form a suitable basis for its empirical analysis nor the information of policy.

What this brief description of how the normative dimension of innovation is addressed in sociologically as well as economically influenced innovation research shows is that these not only differ greatly due to different disciplinary backgrounds, methods, and models, but also primarily conceptualise the negative societal and ethical consequences as a normatively relevant dimensions of innovation. Both in the case of sociological science and technology studies, which follow a constructivist understanding of the inherently normative conditionality and shaping of new technologies, and the economic perspective, which in the mainstream understanding of neoclassical economics assumes a normative equilibrium based on individual preferences and societal compensation, the pivotal point of concern in each case is the theoretical embedding of the normatively relevant societal impacts as well as their practical response – be it through approaches and methods in the sense of technology assessment or through a suitable institutional (policy) set-up (Witt 1996). What is striking, however, is that no comprehensive normative perspective is developed at the level of the concept of innovation itself (Vincent Blok and Lemmens 2015). Against this background, the following section will develop and present a first systematic overview of the different normative dimensions relevant for innovation.

2.2.2.2 Four dimensions of innovation ethics

In the following, a first structuring general overview of the normative dimensions of innovation is developed and presented, on the basis of which existing approaches to dealing with the normative dimensions in innovation research are presented. To this end, a very first systematisation of the normative aspects of innovation is proposed along four dimensions, building on a phase-based understanding of the innovation process (R. G. Cooper 1990, 1983, 1986; Verworn and Herstatt 2000). To the knowledge of the author, as of yet no comprehensive overview or systematization of the relation of innovation and normative implications exists. Therefore, it is important to note that the goal of this chapter is to provide a first general overview of normative aspects of innovation. As a consequence, such a broad scope remains at a more general level and is not apt for highlighting the diversity of manifold specific questions at a micro level. In other words, the aim of this section is to contribute to the emergence of a new field of innovation ethics by providing a first step towards its systematization. My attempt of a first structuring overview is thus by no means meant to be conclusive.

For the proposal of an initial systematisation of the different normative-ethical dimensions of innovation, the structure of a circle was chosen for three reasons. Firstly, it is intended to take into account the increasing dynamisation of value creation and innovation processes, according to which these can no longer be understood in a linear way and along simple push and pull logics. And secondly, the circle illustrates how the different effects of innovation in terms of outcomes and impact contribute to the background conditions for (further) innovation. In other words, innovations always change their own framework conditions – but not always in a deliberate and positive way. And finally, it helps to illustrate that a comprehensive innovation-ethical inquiry can rarely be exhaustive with regard to a single dimension, but should rather be carried with due regard to relevant questions of other dimensions of the circle in order to understand interdependencies and conditional relationships.

Against this background, the four basic dimensions of the circle of innovations ethics are: background conditions, innovation process, innovation output, (side-)effects and externalities.

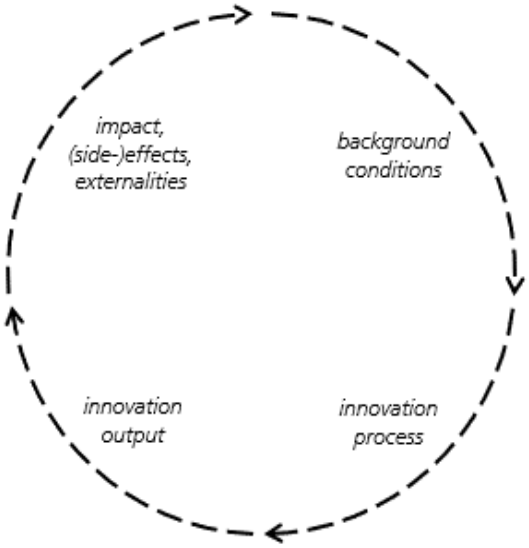

Illustration 1: the circle of innovation ethics

In the following, the introduction and presentation of each of the four dimensions of the innovation ethics circle as a first proposal for systematising a new field of innovation ethics provides not only a brief description of the corresponding normative questions, but also proceeds by offering a brief summary of relevant existing research approaches.

Dimension[4]	background conditions	innovation process	innovation output	impact, (side-)effects, externalities
key question	*Why, when, and what for should we engage in innovation?*	*How should we innovate?*	*What is a good innovation?*	*What are acceptable effects of innovation?*
neighbouring research fields	*political philosophy. political economy, social sciences, normative economics*	*research ethics, business ethics, organizational ethics, science and technology studies*	*domain ethics, applied ethics, normative economics, legal studies, public policy*	*technology assessment, political philosophy*

[4] Note that this table only aims at providing an overview over normative relevant dimensions of innovation. There are of course many more aspects and disciplines important for the study of innovation which are not part of this depiction.

2.2.2.2.1 Normative dimension I: Background conditions

The first dimension, on which normative questions relevant to innovation may arise, refers to the general context and societal background in which innovations are to be developed and used. The scope can be defined in different ways, i.e., both the global context as well as the geographically, culturally, or politically narrow context can be considered. In general terms, all normatively relevant aspects can be addressed that visibly and explicitly – but also implicitly and previously undiscovered – represent preconditions for the development and subsequent use of a particular innovation at hand. This can include, for example, questions about whether an institutional framework (global, international, national, regional) is in place which allows to develop and use innovations in a normatively desirable way. The institutional framework must not only include directly relevant aspects such as questions of (technical) standardisation, sufficient regulation, or processes of public information, but also aspects of the broader institutional framework: Is the innovation embedded in a sufficiently legitimised and democratic legal system? Is (international) tax legislation sufficient and effective to distribute potential profits fairly in society? In a certain sense, such questions can also be linked to J. Rawls' concept of background justice (Rawls 2005, 1958), and it may be built upon in further research for the context of innovation.

Further normatively relevant questions at the level of background conditions can, for example, address the distribution of power: Which actors are able to exploit new knowledge or a certain new technological potential at all? Is there fair competition with other possible ways of utilising the same knowledge or technological potential, which perhaps promise even greater added value for society, but for certain reasons relevant actors are not in a position to exploit it? Of which kind is the competitive relationship for the use of new knowledge and new technologies at all? Are all opportunities of new knowledge or new technologies that promise added value to society really exploited? What should (not) be done so that a technology can be used in the most societally effective way? Along these lines, even the question that can be described as counterfactual to innovation is relevant: What is not invented?

In the context of AI-based technologies, relevant questions concern, for example, the distribution of possibilities to develop and use complex algorithms, models, and systems in the first place, insofar as this requires specialised capacities in terms of data availability, quality, computing capacities, as well as knowledge and skills, which are primarily in the hands of a few global corporations. This raises the question of the extent to which this leads to the technological potential only being used in certain

sectors or for certain fields of application. In order to achieve the highest possible societal impact, however, broad integration into diverse domain knowledge and usage by heterogeneous actors is necessary, for which critical background conditions are missing, need to be questioned, or changed.

The examination of such questions has so far played a rather subordinate role in the context of the assessment of innovation and consequently only a few overlaps with established innovation research exist today. Nevertheless, there are a few examples of studies that deal with comparable normative questions in the context of innovation.

With regard to the question of justice in the diffusion of innovation for example, Buchanan et al. (2011) extend the view not only to the distribution of the products of innovation but also efforts to influence which innovations occur. Justice in innovation thus requires a pro-active stance to utilize the potential of innovation for promoting justice. Innovation ethics in this sense also means developing a concept of justice suitable for the context of innovation (A. Buchanan, Cole, and Keohane 2011). Similarly, Ziegler (2015) explores a concept of just innovation based on Rawls' theory of justice as fairness. Likewise, Woodhouse and Sarewitz (2007) examine the relationship between innovation and social inequality based on the observation that "new techno-scientific capacities introduced into a non-egalitarian civilization will tend to disproportionately benefit the affluent and powerful" (Woodhouse and Sarewitz 2007, 142). Consequently, institutional changes to the science-policy nexus are needed that not only enable the targeted use of innovation to address social inequality, but also guarantee a more equitable distribution of the costs and benefits of innovation. In this sense, social justice should become a primary consideration in deliberations about science and technology policies (Woodhouse and Sarewitz 2007).

Another, so far rather disconnected strand of existing research, which can be classified under the first normative dimension of innovation ethics, deals with the complex relationship between innovation and international development (Papaioannou 2011; Voegtlin and Scherer 2017; Abolhasani et al. 2014; Ludwig and Macnaghten 2020). In this context, Papaioannou (2011) combines innovation research approaches with global justice considerations and argues for a necessary reform of the current intellectual property rights regimes to enable "alternative incentives for successful generation and application of new knowledge" in terms of minimal global justice requirements (2011, 321). Against a similar background of international development, Abolhasani et al. (2014) develop a framework of a justice-oriented innovation system using a grounded theory approach. Drawing on the concept of responsible innovation,

Ludwig and Macnaghten (2020) develop an account of just innovation that aims in particular to integrate the knowledge of traditional communities into innovation processes, especially against an international and global background. In this sense, they reflect on how "underlying political issues such as debates about decolonization, self-determination, and the conditions for change" can be taken into account as prerequisites and framework conditions for change within a justice-oriented conceptualisation of innovation (Ludwig and Macnaghten 2020, 38). Finally, Voegtlin and Scherer (2017) examine deliberative governance schemes in the sense of voluntary soft-law regulations that are suitable for enabling good innovation in the sense of sustainable development in a globalised world.

In the context of normative reflections on the background conditions of innovation, so-called innovation for good (Luciano Floridi et al. 2020; Cowls et al. 2021), inclusive (Schillo and Robinson 2017; Bryden and Gezelius 2017), or sometimes social innovation approaches (Howaldt and Schwarz 2010; Choi and Majumdar 2015; F. Moulaert, MacCallum, and Hillier 2013; Frank Moulaert et al. 2017) may also be considered, insofar as they treat certain normatively relevant shortcomings as objects of innovation.

Overall, the first dimension of the innovation ethics circle might be understood against the backdrop of R. Nelson's influential reflections on the very uneven technological progress that innovation seems to have led to so far (Nelson 1977, 2011). Insofar as Nelson focuses on the systemic and structural backgrounds that have created the particular orientation of innovation systems given t^^oday and points to effective governance measures and policies, he addresses the normative direction of innovation in the sense of the relevant background conditions.

2.2.2.2.2 Normative dimension II: Innovation process

The second level of the innovation ethics circle as presented here concerns the innovation process, i.e., the normative aspects of developing and producing new knowledge and innovations. Which perspectives, needs, and values find their way into the development of new technologies, and which seem to be implicit and structurally excluded from the outset? What normative implications are contained in the systemic framework conditions for doing innovation, be it at the national level of innovation systems, at the level of organisational systems, or of standards and norms for technology-specific development processes? In addition to normatively relevant

questions of the practical design of innovation processes, organisational questions also play a role, e.g., the influence of organisational culture on innovation (cf. Chen, Sawyers, and Williams 1997). The second level of innovation ethics also includes the normative dimension of terms and concepts that are crucial for the innovation process: How are criteria such as truth, novelty, better, or more even defined (J. C. Schmidt 2007)? What (normative) criteria apply to the definition of problems and initial questions relevant to innovation on the one hand and to legitimate solutions on the other? Such questions come to the fore especially in increasingly inter- and transdisciplinary innovation processes, since here different concepts, languages, and (implicit) value systems collide and are being negotiated simultaneously. Finally, questions of collaboration and cooperation may also be raised, insofar as ever more complex innovation processes are usually developed in network-like structures and collaborations. Against this background, ethical questions of collaboration between different actors such as organisations or individuals are relevant on a macro level with regard to power relations or the negotiation of different interests, as well as on a micro level with regard to individual collaboration, including psychological and behavioural aspects. But also issues concerning (implicit) assumptions regarding potential target and user groups are part of the second normative dimension of innovation, insofar as a more or less conceptualised idea of users or customers and their preferences and values is necessary in the course of the innovation process. However, this can raise important questions in terms of biases and prejudices towards people or a particular group.

The concept of responsible innovation has made a significant contribution, particularly with regard to addressing normative aspects of the design of innovation processes. As a more detailed introduction of the concept is given in chapter 3, the importance of the concept for the shift of the scientific view on the innovation process is only briefly mentioned at this point. Although there are different interpretations of responsible innovation in terms of a focus on the process or product level (B. J. Koops et al. 2015), the process-oriented operationalisation of the concept by R. Owen et al. (Richard Owen et al. 2013a; R. Owen, Macnaghten, and Stilgoe 2012) has become of particular importance for research. They define four crucial dimensions of responsible innovation that aim to shape the innovation process in such a way that innovations can contribute to a desirable future. In this sense, the four dimensions focus on the early anticipation of possible impacts, the reflexivity of the actors and institutions involved, the targeted opening up to and integration of previously neglected perspectives and knowledge bases, and the capacity to align innovation processes with the different

perspectives and values of diverse stakeholders (a more detailed overview of responsible innovation can be found in Chapter 3). Based on the concept of responsible innovation, a multitude of methodological approaches, conceptual and empirical studies have emerged that deal with the normative challenges of the innovation process. Not only are different methodological approaches to adapting innovation processes examined (Long et al. 2020), but also institutional aspects (Richard Owen et al. 2021) and the organisational level (Christensen et al. 2020) are taken into account. Another conceptualisation of responsible innovation that focuses on the practice of innovation is offered by the approach of innovation as ethos by V. Blok (2019b) in the sense of "a concept of action-based responsible management of corporate innovation". Similarly, virtue ethics approaches link responsible innovation to professional ethics ideas and seek to operationalise the concept of practical knowledge as a key virtue for innovators (Grinbaum and Groves 2013; Steen, Sand, and Van de Poel 2021; Sand 2018).

In a similar way (Simon 2016; van den Hoven 2013), value-sensitive design focuses on the normative dimensions of innovation processes and the development of new technologies (Friedman and Henry 2019; Zhu et al. 2018). In particular, value-sensitive design focuses on the values of the people involved that are implicitly inscribed in the development of a new product in the course of the innovation process, trying to make them visible, reflect them and thus also make them an object of debate and design - and not to integrate them implicitly and passively. This idea is taken up by so-called "ethics by design" approaches, which place the addressing and explicit consideration of normative challenges at the level of the design of products and solutions and in this sense aim to produce ethical innovations (Dignum et al. 2018; Van de Poel 2001; D'Aquin et al. 2018; Bonnemains, Saurel, and Tessier 2018). Other more practical approaches include considerations that aim to identify normative aspects via the explicit inclusion of ethics experts in innovation processes (van Wynsberghe and Robbins 2014) or in the sense of up-stream engagement (Wilsdon and Willis 2004) or mid-stream modulation (E. Fisher, Mahajan, and Mitcham 2006) through the early inclusion of diverse societal perspectives and to bring them to a broader discussion. So-called professional ethics (Abbott 1983) are also a way of enabling normative reflection at the level of the innovation process and determining certain ethical standards for specific professions such as physicians and engineers. For example, in the context of AI technologies comparable approaches are currently being discussed. From a professional ethics perspective, for instance, P. Boddington (2017) presented the proposal of a code of conduct for the development of AI systems. The Bertelsmann Foundation (2018) introduced a similar proposal for the German

context. B. Mittelstadt (2019), on the other hand, expressed concerns about the effectiveness of a professional ethics approach for the context of AI. McLennan et al. (2020), meanwhile, take up considerations on the inclusion of ethics experts and develop an embedded ethics approach for the development of AI systems.

The second level of the innovation ethics circle may thus draw on various existing approaches to address normative aspects at the level of the innovation process. Nevertheless, it must be noted that these each highlight specific aspects in the context of the innovation process (process design, institutional framework conditions, professional ethics) and that no systematic overview of normative questions and approaches with regard to the innovation process is available to date.

2.2.2.2.3 Normative dimension III: Innovation outputs

The third level of the innovation ethics circle concerns the result of an innovation process, or the innovation itself, so to speak. This is to be understood in the sense of a broad concept of innovation which, in addition to new products, services, and business models, also includes new forms of organisation, changed behaviour, or new social practices. Normatively relevant questions concern both conceptual background assumptions, e.g., with regard to definitions and criteria for what can legitimately claim to be an innovation result at all, and specific ethical questions in the context of the individual innovation and its application. However, in contrast to other dimensions, this is less about the potential effects of introducing and using an innovation than about ethical aspects of a technology, product, or business model itself. What ethical questions arise in the (commercial) use of a technology per se or its marketization through a particular business model? Who should or can benefit from an innovation, who cannot?

Prominent examples in this field include innovations in the field of biotechnology, e.g. (heritable) genome editing, where complex ethical questions already emerge in relation to the technology itself (National Academy of Sciences; National Academy of Medicine; National Academy of Engineering 2017), which have led, among other things, to calls for a moratorium in order to determine ethical guidelines and limits for this technology and innovations based on it (Lander et al. 2019). Similarly, in the field of lethal autonomous weapon systems (Leveringhaus 2016), for example, the general development and use of the technology and innovations in this field is critically discussed (Sharkey 2010). This also implies questions for actors such as companies

conducting research and seeking to develop innovations in this field in order to remain competitive. From an innovation ethics perspective, it could be argued that companies should refrain from innovating in this field and focus on other, possibly new technologies or fields of application (Byrne 2018).

Characteristic for the description of the third level of innovation ethics is, however, that not only technology-, product- or application-specific questions in the sense of domain ethics (Maring 2016) are raised, but that in innovation these coincide, at least most often, with entrepreneurial and/or strategic decisions for a profit-oriented commercialisation or any other market introduction. In other words, in innovation, normative dimensions of research, development, and technology come together with economic and business ethical aspects. Interestingly, however, it is precisely this interface that has so far been rather neglected in the academic debate, possibly due to the complexity of the normative-ethical interrelationships. It seems clear, however, that approaches that examine innovations solely from the perspective of technology ethics lose sight of important dimensions of the economic and entrepreneurial embedding of innovation, and vice versa. With a few exceptions, which do not build on or even refer to each other (Hanekamp and Wütscher 2007; Fassin 2000; Clausen and Loew 2009; di Norcia 1994; Pellé 2017; Preuss 2011; Rexhepi, Kurtishi, and Bexheti 2013), or focus on specific sub-topics such as organisational innovation (Brusoni and Vaccaro 2017) or entrepreneurship (Miles, Munilla, and Covin 2004), there is no established research strand that deals with the topic of innovation from a business or corporate ethics perspective. Yet various ethical aspects of the development and introduction of business strategies (F. Hansen and Smith 2006; Spiller 2000) and specific business models (Lissack and Richardson 2003), which introduce innovations into the market and thus lead to a societal use, are of crucial importance for the normative assessment of innovation. The example of the increasing influence of digital technologies and new possibilities of data analysis on new business models (Ciampi et al. 2021; Loebbecke and Picot 2015) shows that a normative evaluation of this form of commercial use of new technological means is necessary, but not trivial. In this context, major incidents such as the Facebook scandal, in which the company Cambridge Analytica leveraged user data to deliberately manipulate the 2016 presidential election in the United States of America through tools such as microtargeting, might be a case in point. Although this represents an exceptional example, such developments should be understood against the backdrop of more general trends, such as data-based business models, especially in the online advertising sector or in the context of social media, which call for an ethical assessment of such innovations. In addition to the opportunities for

targeting (Ward 2018; Borgesius et al. 2018), the tactical exploitation of so-called dark patterns of human psychology through instruments of user experience design (Gray et al. 2018) and the strategic influencing or even manipulation of behavioural nudges on the basis of Big Data (Yeung 2017; Häußermann 2019) also come to mind. In short, increasing excesses of a so-called attention economy (T. Wu 2016), which harvests our attention by means of new technological possibilities and scientific findings, show that precisely the manner of introduction and commercial use are important factors in the normative evaluation of innovation.

Moreover, so-called dual use issues are relevant in this context, shedding light on the ethical aspects of new knowledge and technologies that can be used for good and harmless as well as for harmful and malicious purposes (Pustovit and Williams 2010; S. Miller 2018). Originally discussed in particular against the background of technologies that can be used for both civilian and military aims, questions of dual use technologies are gaining new importance, not least in the context of digital and biotechnologies. One thinks, for example, of the possibilities of digital technologies for networking and communication, which are used both by oppressed minorities and terrorist perpetrators of violence, or surveillance systems used by authoritarian political systems. Not least, dual use technologies are therefore also an important subject of regulation with regard to security and export issues (European Commission 2009).

On the output level of innovation ethics, the strategic relationship to regulation may also be addressed, in so far as the design of an innovation or a business model is also specifically adjusted to existing legal framework conditions (Coad et al. 2021). However, this also entails the risk that innovations not only attempt to circumvent or delay regulation (Luciano Floridi 2019), but also exploit it in a deliberate manner or even influence it in order to promote innovations that are profitable in the short term but costly and even harmful to society in the long term (Soete 2013).

Finally, at the output level, questions and approaches that make certain normative goals the purpose of their innovation are of course also relevant. Insofar as the possible impacts of the innovations are not dealt with here, it is primarily a question of whether the design of the product or business model is aligned with certain normative goals. To the extent that every innovation implicitly possesses a normative orientation, the discussion of purpose at this level of innovation ethics can be relevant for every kind of innovation, even if the goal is expressed in financial values such as increasing sales or profits. In contrast, approaches such as social innovation (Choi and Majumdar 2015; F. Moulaert, MacCallum, and Hillier 2013; Majumdar, Guha, and Marakkath 2015), inclusive innovation (Chataway, Hanlin, and Kaplinsky 2014; Schillo and

Robinson 2017; Altenburg 2009), sustainable or green innovation (Pesch 2012; Schiederig, Tietze, and Herstatt 2012; Adams et al. 2016), or innovation for social good (Cowls et al. 2021; Luciano Floridi et al. 2020) seek to bring an explicit normative orientation of innovation to the fore and thereby align innovation with specific social goals.

In summary, it can be said that the third dimension of the innovation ethics circle, the level of the outcome of an innovation process, brings together all those normative aspects that are already implicitly or explicitly raised by the design of a product, service, or business model even before the potential effects of its introduction and use. Thus, at this level, questions of the ethical design of new technologies are brought together with aspects of economic and business ethics, insofar as a significant share of today's innovations is introduced into society via commercial markets. With a few exceptions, this nexus seems to have remained rather underexplored in innovation ethics.

2.2.2.2.4 Normative dimension IV: (Side-)effects and externalities of innovation

The fourth level of the innovation ethics circle comprises the normative dimensions of the potential societal impacts of innovation. This category includes all effects that are directly related to the development, introduction, and use of an innovation and are either known and even intended components of the innovation or an accepted effect, as well as so-called unintended consequences that may not have been possible to know and only occur at a later point in time. The fourth level of innovation ethics thus not only encompasses a whole range of different normative dimensions of the complex effects that innovations can cause, it is also the most widely studied field of innovation ethics so far, especially from the perspective of science and technology studies as well as from an economic perspective (cf. chapter 2.2.1). Although from a normative point of view it may initially appear to make sense to also consider and systematise the various positive effects and externalities of innovation (Mulgan 2016), the various potential negative effects constitute the main focus of research. Although negative consequences are often referred to as unintended and undesirable side effects, it is not easy to determine whether and when potential side effects were really not known, could not have been known, or were perhaps part of the system after all ("features, not bugs") and may even have remained intentionally unanticipated to avoid accountability (Coad et al. 2021; Stilgoe 2020). While the sociological discussion tends to refer to (unintended) side effects and consequences (Grunwald 2009; Bechmann et al. 2007),

in economic research potential societal impacts that are not explicitly part of the cost-benefit calculation of the innovation are summarised under the term negative externalities (Witt 1996; Komlos 2016). Roughly summarised, a distinction can be made between pecuniary or distributive and technological side effects or externalities (Binder and Witt 2019)[5]. In principle, the identification, investigation, and evaluation of potential impacts are faced with the challenge that, due to the epistemic quality of innovation as novelty, they may not be recognisable ex ante and may occur at any point in time.

In the sense of Schumpeter's normative foundation, which is still authoritative at least for the economic perspective on innovation, according to which in the aggregate "innovation raises the living standard of the masses" negative pecuniary externalities in the sense of an unequal distribution of welfare effects always occur in society (Binder and Witt 2019; Witt 1996; Komlos 2016). Therefore, the question of how the different welfare effects or the costs and gains of an innovation are distributed among the individual members of a society and how this distribution is normatively justified is crucial. However, the very conceptualisation of costs and gains in a complex economy is anything but trivial (Schubert 2009), not least as they may change against the background of different points in time of an evaluation of the effects. Welfare effects can not only be expressed in concrete monetary values, but also, for example, in terms of the loss of a job, a devaluation of human capital, dislocation, a change or even loss of social status, or simply in terms of uncertainty, anxiety (Schubert 2013, 2015a). Such negative pecuniary externalities can occur both as welfare effects of individual members of society (persons, companies, sectors) and as harm to the economy as a whole. As Binder and Witt (2019, 351) put it: „The reason is that innovations selectively improve the competitive position of single firms or single industries and induce substitution processes at the expense of other firms or industries competing for the same customers' spending. As a consequence of such "pecuniary" externalities, specific investments – made before an innovation was introduced – may be devalued or even lost. Capital owners may face losses of expected returns. Labor may face being laid off and forego expected returns on human capital investments when forced to accept employment elsewhere. In short, innovations often cause massive interpersonal redistribution of welfare."

[5] In line with the interdisciplinary approach of the thesis, the terms side effects and externalities are used interchangeably in the following.

The basic question is therefore how the differing and necessarily unequal societal distribution of the benefits and burdens of an innovation is normatively justified, i.e., which normative criterion is used for this. Authors such as Witt (1996), Binder and Witt (2019) or Schubert (2012, 2013, 2015a) argue from an evolutionary economics perspective that a recourse to welfare economics concepts of mainstream or neoclassical economics based on the fulfilment of individual revealed preferences seems problematic and indefensible from both a theoretical and a practical point of view (see above). Even though the economic perspective is essential for the general understanding of innovation (see 2.1) and the importance of negative externalities is well recognised, there are only few contributions on how to deal with this normative question from an economic perspective. Only a few authors address the normative challenges posed by innovation from an economic perspective in terms of the foundations and concepts of economic theory: from an evolutionary economics perspective (Witt 1996; Binder and Witt 2019; Schubert 2013, 2012), with a focus on policy (Schubert 2015a; Soete 2019), or adopting an empirical approach to the negative economic effects of innovation (Komlos 2016; Mokyr 2014; Soete 2013). Overall, the normative dimension seems to be almost completely neglected in the economic discussion of innovation (Schubert 2015a) and the (implicit) assumption for dealing with the negative effects from an economic perspective is still strongly based on Schumpeter's formative conceptualisation that innovation in the aggregate and in the long run leads to more welfare and progress. Critically, however, this omits an explicit normative discussion of which deprivations can be justifiably accepted by whom, to what extent, and at what point in time in order to make certain welfare gains possible. In short, it seems appropriate to note that, contrary to the influential significance of economics on innovation, there has not yet been a comprehensive discussion of the normative dimension of negative externalities in the sense of the social costs (Coase 1960) of innovation. Rather, critical challenges posed by innovation for standard economic theories seem to be rather ignored and it is adhered to underlying concepts such as Pareto and Kaldor-Hicks efficiency (see chapter 2.2.2), which assume some form of compensation for unequal distributions in welfare effects. In other words, it is a question of designing an institutional embedding through appropriate framework conditions which imply a societally acceptable distribution of welfare gains and losses, especially with regard to negative (pecuniary) externalities (Witt 1996).

Arguably the best-known and best-researched normative aspect of innovation are all the effects that can be summarised under negative technological externalities. Not only

is the list of different examples of negative impacts of new technologies and products long (see chapter 2.2.1), but also the development of different conceptual approaches and empirical studies is most advanced, especially in the sociologically driven discussion of new technologies. Negative technological externalities encompass the entire range of all possible negative societal effects: from negative effects on the health of individuals such as users or those affected and health damage at the level of society as a whole (public health), negative effects in the sense of individual rights violations or restrictions and effects at the level of political-societal coexistence, or negative effects on so-called public goods such as damage to the environment and the climate. In particular, technology assessment approaches offer important methods for analysing and evaluating the complex societal impacts in terms of technological externalities (Bechmann et al. 2007; Grunwald 2002; Müller 1996; Coates 1982; Decker et al. 2004; Coates 2001; Banta 2009). Originally developed in the USA as an institutional instrument of science-based policy advice on new technologies in the late 1960s and 1970s (Bimber 1996), various methods and concepts developed over time both in a academic context and in forms of institutionalised policy advice. The concept of responsible innovation also has its origins in the emergence of technology assessment (see chapter 3.1). As described in chapter 2.1, in the course of the second half of the 20th century the understanding of innovation, innovation systems, and the processes of knowledge production changed fundamentally towards a dynamic and systemic-oriented perspective, which emphasises different actors as sources of knowledge and particularly the co-productive form of the production of new knowledge and innovation, taking into account societal demands (Gibbons et al. 1994; Nowotny, Scott, and Gibbons 2003; Funtowicz and Ravetz 1993). Against this background, a constructivist understanding of innovation in the sense of a "shaping of technology" became the defining understanding of the relationship between new technologies and societal needs and values in the course of the 1980s (Bijker, Hughes, and Pinch 1987; Bijker and Law 1992). In this sense, the identification, analysis and evaluation of ethical, legal, social aspects or implications (ELSA/ELSI) developed over time into an integral part of science and technology studies and innovation research. Parallel to this, in the course of the 1960s and since then, the profound effects of an increasing mechanisation of modern societies became increasingly visible and perceptible (Beck 1986; Jonas 1979), notably through increasing effects of economic growth on the environment (Meadows et al. 1972) or incidents of catastrophic scale such as the Chernobyl disaster in 1986 or with regard to socio-political effects and implications for democratic societies (Habermas 1968). This led, among other things, to a conceptual advancement of technology assessment from expert-based forms (of policy advice) to

approaches that attempt to take research, development, and innovation processes into account (constructive technology assessment) (J. W. Schot 1992; Rip, Misa, and Schot 1995; J. W. Schot and Rip 1997; Genus and Coles 2005; Genus 2006), or propagate the inclusion and participation of new stakeholders, including civil society actors such as citizens (participatory technology assessment) (Abels and Bora 2004; Hennen 1999; J. Hansen 2006; Durant 1999; Cotton 2014; Joss and Bellucci 2002). Last but not least, the concept of responsible innovation emerged from such considerations (see chapter 3.1).

In addition to more recent participatory approaches, almost all quantitative and qualitative methods of (social) sciences are part of the toolbox of technology assessment in order to identify, analyse, and evaluate impacts on all areas of society. With a focus on ecological effects, the approach of life-cycle assessments (LCA) has recently established itself as an influential standard in the investigation of the societal impacts of technologies (Freudenburg 1986; Renn 1982) (ISO 14044) (International Standard Organisation 2006). However, based on the 3-pillar model of sustainability (Renn et al. 2007), social dimensions are increasingly being integrated into corresponding concepts and methods (UNEP/SETAC Guidelines) (Benoît et al. 2010). So-called social life cycle assessments (SLCA) (R. Wu, Yang, and Chen 2014) then face the challenge of complementing environmental- and biophysical-oriented LCA methods based on validated, mostly quantitative indicators with corresponding and comparable categories in the social domain (Grießhammer et al. 2007, 2006; Lehmann et al. 2013). To this end, first methodological and conceptual approaches were developed (Andreas Jørgensen et al. 2008; Iofrida et al. 2018), most of which aim to quantify social effects as well (Corona et al. 2017; Takeda et al. 2019). The Helmholtz-Centre for Environmental Research (UFZ), for example, has developed an approach that extends the typical product-based analysis to a SLCA of an entire value chain of a region (Siebert, Bezama, O'Keeffe, et al. 2018; Siebert, Bezama, Zeug, et al. 2018). Furthermore, first participatory approaches in the field of SLCA are available, for example for the identification of stakeholders and relevant impact areas (Mathe 2014) or to complement them with qualitative research methods or Multicriteria Decision Analysis (MCDA) (De Luca et al. 2015). Approaches for so-called Life Cycle Sustainability Assessments (LCSA) go one step further, combining the analysis and assessment of ecological impacts via LCAs and social effects via SLCAs with an investigation of the economic sustainability dimension with the help of Life Cycle Costings (LCC) (Finkbeiner et al. 2010; Kloepffer 2008). The advantage and at the same time the major challenge of corresponding assessment models is the

comparability and integration of the different values from the ecological, economic, and social dimensions.

On a more thematic level, Coad et al. (2021) recently identified five different dimensions in which innovation can be harmful and derived four integrated types of harm from innovation for society on an aggregated level. First, they identify *public health risks* in the sense of known health risks as well as the general distribution of benefits and costs, e.g. through opportunity costs of investments in expensive high-tech medicine. A second category of possible harms is the potential influence of innovations on *environmental degradation*. This includes not only direct climate damage in the sense of high CO2 emissions, but also "the collapse of some ecosystems and regular mass extinctions, involving the melting of Arctic and Antarctic ice, thawing of permafrost, destruction of coral reefs, irreversible fragmentation of rainforests, the expansion of deserts, increasing frequencies of droughts, floods and forest fires, and the collapse of insect and vertebrate populations" (Coad et al. 2021, 6). Under the third category, *harm to society*, the authors summarise all those negative consequences that innovations can have on society and a good way of living together. Think of, for example, what impact algorithmic systems can have on discrimination and the fair treatment of (groups of) people, or what impact digital platforms can have on polarised public debates and democratic opinion-forming processes. *Harm to the economy*, the fourth type of harm, encompasses the negative economic effects that innovations can have, for example by leading to monopoly effects on markets or the exploitation of patent and IP rights leading to unequal knowledge distributions that do not efficiently exploit the societal benefits of new technologies.

Biggi and Giuliani (2021) go one step further by using bibliometric techniques to identify thematic focus areas in the academic debate on the negative impacts of innovations. Accordingly, five different thematic clusters can be identified in a first systematic literature review of the research field (Biggi and Giuliani 2021): While the cluster of *work-related consequences of technology acceptance* deals with the adverse effects, in particular of the use of ICT technologies in organisations, e.g. with regard to job satisfaction, work-life balance, and the impact of ICT on the workplace, the rapidly growing field of *unsustainable transitions* encompasses studies that deal with the unintended and harmful effects of certain socio-technical systems on a meso level (e.g., electricity, agro-food, transportation). In particular, the focus is on the unsustainable tendency of such innovations and the risks of so-called rebound effects of innovations that are actually supposed to make a positive contribution to the environment. The third cluster, *innovation & growth downside effects*, focuses on the

impact of innovations on a macro level, for example in the sense of the effects of new financial instruments or effects on employment and economic inequality. The fourth identified thematic area, *dangers of emerging technologies*, is dedicated to the investigation of possible unintended negative consequences of emerging technologies such as nanotechnologies and aims at appropriate governance frameworks. Finally, the last cluster, *open innovation's dark side*, looks at the negative impacts of open innovation strategies as organisational innovation on the performance and survival of firms. Based on a total of 125 studies Biggi and Giuliani (2021) identify these five thematic clusters as well as their respective theoretical antecedents. Overall, the second cluster, *unsustainable transitions*, not only has the largest number of assigned articles, but shows the highest growth in recent years, along with topic area four, *the dangers of emerging technologies.*

In accordance with the theoretical differentiation introduced in this thesis between social science-driven perspectives of innovation research on the one hand and more classical economic approaches on the other (cf. 2.1.2; 2.1.3; 2.2.2.2; 2.2.2.4.), the identified types of harm (Coad et al. 2021) and thematic research strands found in the literature (Biggi and Giuliani 2021) can also be categorised accordingly, insofar as there is one area (*harm to the economy*) and two thematic clusters (*innovation & growth downside effects & open innovation's dark side*) that are explicitly dedicated to the economic dimension. All others can be assigned to sociologically-oriented science and technology studies. Overall, it is striking that in these most recent – and to the author's knowledge only – reviews of the research field on the normative dimensions of innovation, the focus is more or less exclusively on the effects in terms of negative (side) effects and externalities. Even negative pecuniary externalities in the sense of the distribution of welfare effects seem to appear rather marginally. This not only underlines the enormous importance of the normative dimension for the field of innovation ethics, to which a contribution is to be made here. It also shows that the relevance of normative aspects is still primarily located at the level of (non-intended) consequences and is less understood as a normative dimension of the concept of innovation per se. This in turn underlines the relevance of addressing normative-ethical questions in the context of innovation from a more comprehensive perspective. The concept presented here, based on four integrated dimensions of the innovation ethics circle, offers a first proposal for further discussion.

3. Examining the prospects of responsible innovation

This chapter presents a brief introduction into the concept of responsible innovation, originally introduced as „responsible research and innovation (RRI)“, providing the conceptual background of the thesis and the three publications it is based upon.[6] As the field of responsible innovation today encompasses a wide variety of conceptual and empirical studies, practice-oriented publications, different forms of institutionalization and organization, as well as the practical development, test, and use of various methods and tools, the aim of this section is not to provide an exhaustive and concluding introduction to the emergence the concept, and its use and role today. Rather, this section seeks to give the reader a solid understanding of the concept of responsible innovation in order to be able to comprehend the identified research needs that have been addressed in the three published case studies. Therefore, the general introduction of the concept is followed by three sub-sections focussing on the specific aspects of responsible innovation which this thesis focuses on. The three sub-sections thus present the more closely related literature in the field of responsible innovation relevant to each of the three research questions of the individual articles.

[6] When the idea of a concept of responsible innovation was originally developed and introduced it seeked to comprise the whole innovation process and in particular phases of research and development as enabler and early phases of innovation. Following a wide conceptual understanding of innovation (see section 1.1.2), however, one can assume that the notion of research is an inherent part of innovation. Moreover, as the focus of this thesis is not on the particular characteristics of research and its normative and ethical implications, as in the case of research ethics for example, but rather it is interested in the often economical- and business-driven use of innovation in today's societies, the notion of "responsible innovation" is used.

3.1 A brief introduction to the concept of responsible innovation

In dealing with the normative dimensions of innovation, the concept of responsible innovation has established itself as a particularly valuable approach in recent years. Today, responsible innovation not only combines different streams of innovation research, but also increasingly integrates other disciplines and approaches, for example from economic and business ethics or organisational sociology. As a concept which was very much policy-driven in its origins, it is characterised by a peculiar mixture of theoretical, conceptual, and empirical research on the one hand and practical approaches such as innovative methods and tools for implementation on the other. In this sense, responsible innovation has always been not only an interdisciplinary, but above all a transdisciplinary field. At least in part, the concept of responsible innovation was developed in the wake of the Lisbon Treaty in 2009 and as part of the completion of the European Research Area (ERA), gaining increasing attention especially since 2011 onwards (de Saille 2015; Rip 2016; Tancoigne, Randles, and Joly 2016). Responsible innovation has thus become a central element of the European Commission's science, technology, and innovation (STI) policies, particularly through the influence the concept has had on the various research framework programmes, the design of funding streams, and even the formulation of individual calls for proposals (Stirling 2006). While before societal dimensions of innovation were rather conceptualised in terms of a deficit and as the management of possible irrational public resistance (de Saille 2015), the name of corresponding programmes changed from "Science and Society" to "Science in Society" in the context of the FP7 research framework programme (in which responsible innovation experienced an explicit roll-out for the first time), before it finally appeared in the Horizon 2020 framework programme (FP8) as "Science with and for Society“ (SwafS) (R. Owen, Macnaghten, and Stilgoe 2012). How responsible innovation will find its way into the latest research framework programme "Horizon Europe", on the other hand, seems to be an open question at this point in time (Erik Fisher 2020).

At the level of the academic discourse, responsible innovation builds in particular on various concepts of technology assessment (cf. 1.2.2), which in turn have been strongly influenced by debates in the field of bioethics and the ethical questions surrounding nanotechnologies (Grunwald 2012; Burget, Bardone, and Pedaste 2017). Building on a constructivist understanding of innovation, according to which the development of new knowledge and technologies is an inherently social process (Yoshinaka, Clausen, and Hansen 2003; Bijker and Law 1992; Bijker, Hughes, and Pinch 1987) which in addition to scientific and technological inventions necessarily involves social

dimensions and should therefore include institutional frameworks as well as a variety of affected perspectives and actors. Concepts such as constructive (Rip, Misa, and Schot 1995; J. W. Schot and Rip 1997; Genus 2006; Genus and Coles 2005) and participatory technology assessment (Hennen 1999; Joss and Bellucci 2002; Joss 2002; Cotton 2014; Hennen, Petermann, and Scherz 2004; Hennen 2012), along with other influences such as anticipatory governance (Barben et al. 2008; D. H. Guston 2014), midstream modulation (E. Fisher, Mahajan, and Mitcham 2006), or the ethics of responsibility (Lenk and Maring 1993; Jonas 1979), are significant precursors to the concept of responsible innovation (Ribeiro, Smith, and Millar 2017; Burget, Bardone, and Pedaste 2017). The concepts of constructive and participatory technology assessment already point to two core elements of responsible innovation, according to which a responsible shaping of innovation can succeed by opening up to diverse societal perspectives and taking them into account at an early stage, i.e., already during the innovation process. Against this background, the three most relevant definitions of responsible innovation are as follows:

„Responsible Research and Innovation (RRI) refers to the comprehensive approach of proceeding in research and innovation in ways that allow all stakeholders that are involved in the processes of research and innovation at an early stage (A) to obtain relevant knowledge on the consequences of the outcomes of their actions and on the range of options open to them and (B) to effectively evaluate both outcomes and options in terms of societal needs and moral values and (C) to use these considerations (under A and B) as functional requirements for design and development of new research, products and services“ (European Commission 2013, 3).

„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)“ (Rene Von Schomberg 2011, 9).

„Responsible innovation means taking care of the future through collective stewardship of science and innovation in the present“ (Stilgoe, Owen, and Macnaghten 2013, 1570).

Along these lines, responsible innovation can be defined as a participatory governance framework that aims to align new technologies with societal values (Richard Owen et al. 2013b). Through the deliberative inclusion of different stakeholders and, not least, individual members of the public, the aim is not only to anticipate negative

implications and prevent potential harmful effects, but ultimately to make innovation useful for addressing societal needs and solving major societal challenges, in short for societally desirable futures. Through the fundamental participatory-deliberative approach, responsible innovation is based on an understanding of broadly shared responsibility ("co-responsibility") to control innovation and steer it in an ethically acceptable, societally desirable, and sustainable direction (Lubberink 2018). On this generally shared basis, a variety of different conceptualisations exist to date (Burget, Bardone, and Pedaste 2017; Ribeiro, Smith, and Millar 2017), whereby a fundamental distinction can be made between process-oriented and outcome-oriented approaches (B. J. Koops et al. 2015). While the latter assumes a substantial normative orientation of innovation towards certain goals and values and asks for the right impacts of innovation (René Von Schomberg 2013, 2014), a more process-oriented understanding seems to have become more established, particularly in the academic debate, operationalising responsible innovation along the four dimensions of *anticipation, reflexivity, inclusion/deliberation, responsiveness* (Richard Owen et al. 2013a; Stilgoe, Owen, and Macnaghten 2013).

Accordingly, responsible innovation requires to be anticipatory in the sense that innovation should identify both intended as well as possible unintended economic, social, ethical, environmental, or other consequences. While anticipation in this sense means the identification of various trends and current developments in order to recognise possible (negative) externalities as early as possible, it is also accompanied by a more comprehensive examination of the wider environment of innovation, e.g. in the sense of social needs (Lubberink et al. 2017). Closely linked to its roots in different forms of technology assessment, the anticipatory dimension of responsible innovation therefore means, in addition to recognising the necessary uncertainty associated with developing and adopting an innovation, identifying and analysing potential impacts and developing measures and strategies to prevent or mitigate them and strengthen the positive impact. The second dimension of reflexivity aims at the capacities and practices to deal with the necessary confrontation with known framework conditions as well as unknown and even unconscious influences on innovation. This can include reflection on governance mechanisms, motivations, normative assumptions but also areas of ignorance and uncertainties (Richard Owen et al. 2013a). In addition, forms of reflection can find expression both in the implementation of certain methods or standards, the reflective design of innovation processes, or the adaptation of organisational framework conditions and the building of reflective capacities. The third dimension of responsible innovation, inclusion or deliberation, presents, in a sense, a cross-cutting aspect (Burget, Bardone, and Pedaste

2017) insofar as it can also be seen as a tool or response to realise the requirements of anticipation and reflexivity. In short, this dimension determines to „inclusively opening up visions, purposes, questions, and dilemmas to broad, collective deliberation through processes of dialogue, engagement, and debate, inviting and listening to wider perspectives from publics and diverse stakeholders. This allows the introduction of a broad range of perspectives to reframe issues and the identification of areas of potential contestation“ (Richard Owen et al. 2013a, 38). The underlying assumption is thus that the different normative aspects of innovation require the inclusion of different perspectives or even a broader societal exchange in order to bring about a decision. This not only takes into account increasing demands for legitimacy in the context of the governance of science, technology, and innovation (see section 1.1.3), but also facilitates the development of "better" solutions, e.g. in the sense of accepted innovations or new fields of application and business models (Stilgoe, Owen, and Macnaghten 2013). In this sense, the dimension of inclusion or deliberation can be used and operationalised with different objectives, both for normative considerations in the sense of involving directly or indirectly affected stakeholders and for instrumental purposes such as for a better understanding of societal needs or for the capitalization of resources, knowledge, or capital (Lubberink et al. 2017). While the term inclusion primarily emphasises the requirement of involving different stakeholders, deliberation also brings the mode of inclusion and exchange to the fore. In short, this dimension manifests the ideal of inclusion, deliberation, or participation as a central element for the conceptualisation of responsible innovation (Brand and Blok 2019; van Mierlo, Beers, and Hoes 2020; Voegtlin and Scherer 2017). Finally, the fourth dimension, responsiveness, describes the requirement for responsible innovation to provide for the "capacity to change shape or direction in response to stakeholder and public values and changing circumstances" (Stilgoe, Owen, and Macnaghten 2013, 1572). Responsiveness can be achieved or enhanced through a variety of mechanisms and addressed at different levels. Building on the investigation of responsiveness in the context of environmental policy (Pellizzoni 2004), responsiveness can be related both to the systemic level of innovation systems and to the organisational level of governance and culture of actors such as companies, research organisations, or public administration agencies. While in the former case, for example, the complex structure of cooperations and processes can be reflected with regard to their responsiveness to societal challenges, organisational responsiveness refers to issues such as the institutional capacities to recognise and understand new and different forms of knowledge, to absorb it, and to make strategic use of it.

With regard to these four dimensions of responsible innovation, a number of different methods, practices, and tools have been discussed, investigated, developed, and applied, both in terms of a descriptive-analytical analysis and the practical implementation of responsible innovation in research and business (Schuijff and Dijkstra 2020). This includes the development of specific methods and tools (Long et al. 2020), as well as approaches for organisational implementation (Christensen et al. 2020) and institutionalisation (Richard Owen et al. 2021).[7] Key activities related to the different dimensions include foresights (horizon scanning, scenarios), technology and vision assessments, living labs, or stakeholder mapping techniques for anticipation; multi-disciplinary collaboration, embedded ethicists, review processes and standards (ethical audits), codes of conduct, or moratoria for reflexivity; focus groups, user-centred or participatory design, citizens panels and consensus conferences, social labs, or crowdsourcing for inclusion and deliberation; and finally organisational measures (leadership, corporate culture), diversity, rules, adjusted regulation & standards, open innovation & open access, value-sensitive design, or absorptive capacity routines for responsiveness (Stilgoe, Owen, and Macnaghten 2013; Lubberink et al. 2017; Schuijff and Dijkstra 2020).

As responsible innovation is a general approach to aligning innovation with societal needs, approaches can be found today across a variety of academic fields and sectors, including information and communication technologies (Stahl, Timmermans, and Flick 2016; Stahl 2012; Jirotka et al. 2017; Stahl and Coeckelbergh 2016; Stahl 2019; Stahl, Eden, and Jirotka 2013), nanotechnology (Malsch et al. 2015; Krabbenborg and Mulder 2015), neurotechnology (de Jong et al. 2016; Garden et al. 2016, 2019; Shelley-Egan et al. 2018), medicine and health (Kerr, Hill, and Till 2018; Dzobo et al. 2020; Chatfield et al. 2017; Pacifico Silva et al. 2018), food (Long and Blok 2017; Garst et al. 2017; Brand 2020; de Bakker et al. 2014), agriculture (Bruijnis et al. 2015; Gremmen, Blok, and Bovenkerk 2019; Rose and Chilvers 2018), biotechnology (Asveld, Ganzevles, and Osseweijer 2015; Douglas and Stemerding 2013; Shortall, Raman, and Millar 2015), or energy innovation (Correljé et al. 2015; Pellizzone et al. 2017, 2015).

Moreover, not only has a community of researchers and practitioners from industry and policy emerged in the last decade, but there are also a number of central resources for dealing with responsible innovation. In addition to a specifically established

[7] A website, rri-tools.com, has been set up to provide an overview of the different methods and approaches.

academic journal, the "Journal of Responsible Innovation" (David H. Guston et al. 2014), the founding volume "Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society" (Richard Owen, Bessant, and Heintz 2013) and the authoritative series consisting of the anthologies "Responsible Innovation 1" (van den Hoven et al. 2014), "Responsible Innovation 2" (B.-J. Koops et al. 2015) and "Responsible Innovation 3" (Asveld et al. 2017) are particularly important. In addition, the originally strongly European-oriented perspective has increasingly been expanded to include international perspectives on responsible innovation (René von Schomberg and Hankins 2019; Macnaghten et al. 2014).

Finally, a further perspective can be cast on responsible innovation. As responsible innovation aims to contribute to a more comprehensive democratisation of innovation on a systemic level (Braun and Griessler 2018), the ambitions of the concept do not only focus on changing innovation processes in the sense of a participatory governance along the dimensions of anticipation, reflexivity, deliberation, and responsiveness. Rather, it also aims at institutional changes of structural and systemic preconditions of the innovation system in terms of the orientation towards societal needs in the sense of a normative directionality (Nelson 1977, 2011). However, responsible innovation should not itself be subject to a pro-innovation bias and overlook the normative entanglements at the level of the concept of innovation itself (Vincent Blok and Lemmens 2015; Vincent Blok 2020). In this sense, responsible innovation will have to escape the cult of innovation (Brand 2020) and recognise the political dimension of innovation (Delvenne 2017). It is here that this dissertation starts and, based on the normative dimensions of innovation (see section 2.2), contributes to a more complete and effective conception of responsible innovation in terms of a re-politicisation of innovation as a whole (Benoît Godin 2008).

3.2 Three research gaps and nearest neighbours in the literature

The overarching research aim of this dissertation is to develop a more solid and comprehensive understanding of responsible innovation given its current conceptualizations as introduced in section 3.1 above and the different normative dimensions of innovation as described in chapter 2. Against this background the dissertation is based upon the assumption that responsible innovation needs to engage more intensively with neighbouring disciplines and research fields in order to be able to cover more effectively the different normative dimensions of innovation. In other words, the aim is to foster responsible innovation by integrating interdisciplinary perspectives from neighbouring fields. More specifically, I argue that current conceptualizations of and approaches to responsible innovation lack perspectives from normative economics, business ethics, as well as from current research on the social acceptance of new technologies. Even though innovation today is perceived following an economic understanding as well as considered an entrepreneurial endeavour which results to the conceptualization of innovation as commercialized and adopted innovation, approaches of responsible innovation seem to be almost blind with regard to (normative) economic theory, not quite favourable towards established concepts of business ethics, and reluctant to incorporate new findings from innovation acceptance research. In particular, this pertains to the role of participation as key for responsible innovation and the notion and relevance of it in related research strands. Therefore, this dissertation and the three individual publications it is based upon seek to address these three shortcomings of responsible innovation. In three interdisciplinary case studies, of which one is theoretical, one conceptual, and one an empirical study, I combine perspectives, concepts, and findings of a neighbouring field of research with the concept of responsible innovation. As the results will show, each of the different approaches can provide a valuable expansion of responsible innovation. The first article of the dissertation, "Aligning Innovation and Ethics: An Approach to Responsible Innovation Based on Preference Learning", connects responsible innovation with an investigation of the normative implications of innovation from a perspective from normative economic theory. The second publication, "Community-in-the-loop: towards pluralistic value creation in AI, or—why AI needs business ethics", starts from a discussion of the ethics AI as new and emerging key technology and develops an approach to responsible innovation as participatory governance building on an established theory of business ethics. The third study, entitled "Social acceptance of green hydrogen: Building trust through participation", explores the role of participation for the social acceptance of green

hydrogen in Germany and thereby allows to demonstrate the prospects of responsible innovation from the perspective of technology acceptance research. The following three sub-sections introduce each combination of responsible innovation with a different field of research, refer to the existing literature on the specific intersection (if there is any), and point to the benefits of such interdisciplinary examinations. Each of the three sub-sections concludes with the derived research questions the corresponding article attempts to answer.

3.2.1 Responsible innovation and normative economics

The analysis of responsible innovation from the perspective of normative economics is based on two observations and resulting assumptions. First, innovation has an inherently normative dimension that is also relevant from an economic viewpoint on the concept of innovation. Secondly, approaches to responsible innovation have so far mostly been "narrow" (Vincent Blok and Lemmens 2015) in the sense that the normative implications from the perspective of economic theory have hardly been taken into account. However, this appears to omit significant aspects, not least given the fact that an economic conceptualisation of innovation not only determines the general understanding, albeit mostly implicitly, but also plays a decisive role in informing policy and the corresponding design of a governance framework for innovation. It seems all the more astonishing that responsible innovation has so far hardly taken into account perspectives and insights from normative economics or even been discussed against the background of economic theory and its conceptualisations of innovation, apart from rather cursory references to Schumpeter's dictum of innovation as "creative destruction". In addition to the article by Schlaile et al. (2018) in particular, the contribution by M. Hühn (2018) should be mentioned which reconstructs A. Smith's conceptualisation of innovation as an exercise in social construction based on his distinct anthropology and the famous impartial spectator. From this, he then develops an account of responsible innovation informed by the Smithian concept of the entrepreneurial process (Hühn 2018). Notably, Schlaile et al. (2018) argue for an „(re-)conceptualisation of responsible innovation by proposing an evolutionary economic approach“. Based on the explicit assumption of innovation as a complex, collective, and dynamic phenomenon, they develop an “agent-based model that incorporates a multidimensional space of characteristics in which new products or services are represented by more than the mere aspect of price and quality. Instead, innovations are denoted by a large set of characteristics, including also negative or harmful ones” (Schlaile et al. 2018, 7). Building on a comparative discussion of neoclassical economics and evolutionary economics and the implications for an economic conceptualisation of innovation, they argue for a broader and more complex consideration of the role of individuals as consumers in the context of innovation. To this end, they finally present an agent-based simulation model that explores in different scenarios the possibilities and implications of a more comprehensive evaluation of innovation by consumers, i.e., a consideration of negative and harmful effects alongside factors such as quality and price. In a sense, the interdisciplinary study by Schlaile et al. (2018) thus builds a first bridge between responsible innovation on the

one hand and the existing economic debate on the normative dimension of innovation (Komlos 2016; Binder and Witt 2019; Witt 1996), especially from an evolutionary economics perspective (Buenstorf et al. 2013b; Schubert 2012, 2015a, 2013). However, in the context of an economic debate these focus on the discussion of the normative justification of innovation against the background of the unequal distribution of welfare gains and losses and corresponding implications for innovation policy (see chapters 2.1.3 & 2.2.2). For example, Witt (1996) compares different institutional set-ups such as a Pareto progress enforcing regime or a laissez faire regime as well as possibilities of a liability or negligence rule in order to distribute the negative externalities of innovation fairly in society. What the first interdisciplinary bridge between responsible innovation and normative economics by Schlaile et al. (2018) shows in particular is the (normative) importance of individuals as consumers. Insofar as this aspect has been rather neglected in the context of responsible innovation, this offers a suitable interface for a closer link between responsible innovation and normative economics.

Against this background, the first article of this dissertation aims at an extension of responsible innovation through a normative economic theory perspective on the normative implications of innovation with a particular focus on the role of individuals. Therefore, the following research questions will be addressed in the context of a *theoretical case study*: (1) What are the implications of the concept of innovation for normative economics and how can a theoretically grounded normative conceptualisation of innovation from the perspective of economics look like? (2) What are the implications of such a normatively justified conceptualisation of innovation from the perspective of economics for responsible innovation, both on a theoretical and practical level, with special consideration of the role of participation?

3.2.2 Responsible innovation and business ethics

With the growing establishment of responsible innovation and the increased engagement with the concept both on a scientific-theoretical level and with regard to a practical implementation of different operationalisations and practice-oriented approaches, the interest in the framework conditions, opportunities, and barriers of responsible innovation in an entrepreneurial and industrial context increased as well. Although the concept originally emerged from policy-driven and academic debates (see section 3.1), the need to address the business level of innovation was recognised early on. This has given rise to a rapidly growing strand of research – perhaps one could even say that this is currently the most active field of research in the area of responsible innovation – which can be summarised under the title "responsible innovation in industry" (Long et al. 2020; Long, Iñigo, and Blok 2020; Fraaije and Flipse 2020; Lubberink et al. 2017; Lubberink 2018; van de Poel et al. 2017; Dreyer et al. 2017; Martinuzzi et al. 2018; van de Poel et al. 2020; Yaghmaei and Poel 2020; V. Blok, Hoffmans, and Wubben 2015; Pellé 2017; Flipse 2013; Timmermans et al. 2017; Iatridis and Schroeder 2016; Auer and Jarmai 2017; Garst et al. 2017). What is striking, however, is that the explicit consideration of business ethics in the sense of classical theories and established concepts is rarely discussed in the context of responsible innovation in industry. This may be due to the fact that responsible innovation is often based on theories of science and technology studies and technology assessment rather than on business ethics (see also sections 3.1, 2.1.3, 2.2.2). A few exceptions to this are the more recent contributions by Brand (2020), Brand and Blok (2019), Brand, Blok, and Verweij (2020), und Blok (2019a). They focus in particular on the notion of "deliberative engagement as a central governance mechanism in responsible innovation" (Brand and Blok 2019, 5) and highlight the different tensions that this aspect leads to in the context of market-based economy and corporate surroundings. The introduction of a governance model for innovation geared towards stakeholder engagement (Silva et al. 2019; Noorman, Swierstra, and Zandbergen 2017) and deliberation may not only give rise to conflicts in a commercial context with regard to negative effects on the innovation capacity of firms, problems of knowledge sharing, the loss of competitive advantage, and not least to conflicts with existing corporate governance structures (Brand and Blok 2019), but also points to complex power relations and potential or even likely conflicts between different interests and values of the stakeholders involved. In this sense, some of the authors argue for an explicit recognition of divergent or even opposing values and for a move away from consensus-oriented models of deliberation and stakeholder engagement. Rather, they

argue, an agonistic form of deliberation seems more appropriate (Brand, Blok, and Verweij 2020; Vincent Blok 2019a) in order to be able to meaningfully adhere to the ideal of deliberative engagement as a governance mechanism for responsible innovation (Brand and Blok 2019). In any case, a closer analysis of the micro-level of participation as a governance mechanism and the ethical implications that arise from it, as ethics of stakeholder engagement or ethics of participation (Vincent Blok 2019a), seems warranted (cf. Häußermann et al. 2021). The few existing contributions that conceptualise responsible innovation against the background of established theories and concepts of business and corporate ethics must also be understood in this context. Brand and Blok (2019), for example, discuss the ramifications of responsible innovation's participation requirements in contrast with Heath's market failures approach to business ethics (Heath 2006, 2014), stakeholder theory (Freeman et al. 2010; Freeman, Wicks, and Parmar 2004; Phillips 1997; Harrison and Wicks 2013) and political Corporate Social Responsibility (CSR) (Palazzo and Scherer 2006; Scherer and Palazzo 2007, 2011) in order to derive lessons for responsible innovation. Voegtlin and Scherer (2017), on the other hand, develop their own account of responsible innovation in a global context of sustainable development using political CSR as a general framework and only selectively referring to the concept of responsible innovation. Beyond that, the emerging field of responsible innovation in industry regularly refers to both research and practices of corporate social responsibility, mostly, however, without explicitly building on a more profound theory of business ethics (Martinuzzi et al. 2018; Dreyer et al. 2017; Iatridis and Schroeder 2016; Vincent Blok 2019a; van de Poel et al. 2017; Chatfield et al. 2017; Garst et al. 2017; Hemphill 2016).[8] It can therefore be observed that so far only a few exceptions discuss responsible innovation from a business ethics perspective and in particular with a focus on the participatory claim of a governance framework based on deliberative engagement. Accordingly, voices point out that this ideal seems to be problematic in commercial contexts (Noorman, Swierstra, and Zandbergen 2017), cannot adequately reflect ethical dimensions and value conflicts (Brand, Blok, and Verweij 2020; Vincent Blok 2019a), and responsible innovation thus seems to be confined to an academic perspective (Silva et al. 2019; Dreyer et al. 2017). As van de Poel et al. (2020) highlight,

[8] What is also interesting to note is that from the perspective of research on CSR too, the area of innovation is only rarely discussed, or at least not systematically (Martinez-Conesa, Soto-Acosta, and Palacios-Manzano 2017; Liu and Xu 2014; Rexhepi, Kurtishi, and Bexheti 2013; Ferauge 2012; Preuss 2011; Clausen and Loew 2009; Groves et al. 2011; Louch, Idowu, and Fllho 2010; Lins, Servaes, and Tamayo 2017; Gallego-Álvarez, Prado-Lorenzo, and García-Sánchez 2011; Saeidi et al. 2015).

this calls for, amongst other things, further strategizing for stakeholder engagement and placing values at center stage.

Against this background, the second paper of this dissertation follows first attempts to connect responsible innovation more closely to established theories of business ethics, in particular with regard to the central governance mechanism of deliberative engagement. In order to contribute to further strategize stakeholder engagement and place values center stage (van de Poel et al. 2020) as well as to be able to adjust responsible innovation to the competitive environment of markets and corporate requirements (Silva et al. 2019; Dreyer et al. 2017), it provides a *conceptual case study* developing an participatory and deliberative governance mechanism building on the established theory of business ethics called order ethics. The research question of this article therefore asks: (1) How can a governance framework based on participation and deliberation be developed on the basis of order ethics as an established theory of business ethics? (2) How can specific normative-ethical issues of an emerging technology like AI be conceptualised as an entrepreneurial challenge to which the participatory governance framework provides a response in the sense of responsible innovation?

3.2.3 Responsible innovation and social acceptance

Albeit with different emphases and interpretations, the social acceptance of innovations played an important role throughout the history of the concept. At times, it was rather the adopter and user than the inventor who was considered the innovator (Benoît Godin 2008). In particular with an growing sociological understanding of innovation, the diffusion and adoption of innovation moved into the focus of attention (Rogers 1962).

Last but not least, the complex social processes responsible for the social acceptance and success of an innovation became part of the academic debate. The undisputed and fundamental economic benefits that innovation has brought in recent centuries have not only led to an instrumental understanding of innovation as growth policy and an entrenched pro-innovation bias, but have also resulted in resistance to innovation or technological inertia being seen as something temporary that will inevitably be overcome and can even only be explained by irrationality (Mokyr 2000; Juma 2016). However, the normative complexity of innovation and the necessary unequal societal distribution of welfare gains and losses demonstrate that acceptance is a complex social process for the integration of new technologies and innovations into societal systems and institutional frameworks. The debate and the struggle for social acceptance thus express nothing other than "tensions between the need to innovate and the pressure to maintain continuity, social order, and stability" (Juma 2016, 5). In this sense, technological inertia is not only an expression of social negotiation processes to balance advantages and disadvantages, opportunities and risks and their societal distribution, but also an important challenge for democracies in times of technological change. Insofar as there is the risk of a "serious inconsistency between democracy and continuous innovation", so that "democratic decision processes do not maximize the long-term economic welfare of economies" (Mokyr 2000, 65), the management of "the interactions between change and necessary social stability remains one of the most critical functions of government" (Juma 2016, 6) – especially of democratic societies.

Against this background, it is hardly surprising that the academic study of the acceptance of innovations and new technologies is an established and widely diversified interdisciplinary field of research. Especially in the field of information technologies, the technology acceptance model (TAM) introduced by Davis (1986) is influential in explaining perceived usefulness and usage intentions in terms of social influence and cognitive instrumental processes. While they focused on the use of

technologies in a working environment (Davis 1989; Davis, Bagozzi, and Warshaw 1989), its continuous elaboration and extension aimed, for example, at a better understanding of perceived usefulness through external factors such as social influence or subjective norms (Venkatesh and Davis 2000; Venkatesh et al. 2003; Marangunić and Granić 2015). In addition to the field of information technologies, a weighty strand of scientific debate on acceptability has been established in the context of renewable energy innovation (Boudet 2019; Devine-Wright et al. 2017; Gupta, Fischer, and Frewer 2012; Upham, Oltra, and Boso 2015; Wüstenhagen, Wolsink, and Bürer 2007). The different definitions, disciplinary approaches, and diverse research methods (Upham, Oltra, and Boso 2015; Ricci, Bellaby, and Flynn 2008; Ruiz, Marrero, and Hernández 2018) share a common interest in understanding and explaining the different internal factors at the level of the individual acceptance subject as well as external factors at the level of the acceptance object and the wider societal and institutional context. In this regard, participation plays an increasingly important role (Holstenkamp and Radtke 2018) as social acceptance is not solely dependent on the properties of a given technology, but is likewise influenced by the interactions and dialogue among various stakeholders. As the promotion of acceptance has always been a central objective of the concept of responsible innovation (Ribeiro, Smith, and Millar 2017; René Von Schomberg 2012; Richard Owen et al. 2013a), the consideration of acceptance from the perspective of responsible innovation seems to be obvious. It is therefore somewhat surprising that the intersection of acceptance research and responsible innovation has only been explored in very few cases (Correljé et al. 2015; Koirala, van Oost, and van der Windt 2018; Pellizzone et al. 2015, 2017). Whereas Correljé et al. (2015), for example, develop an approach to responsible innovation in energy projects with a focus on stakeholder interactions based on the concept of value sensitive design, Koirala et al. (2018) operationalise the four dimensions of responsible innovation based on Owen et al. (2013a) for the field of community energy storage. Pellizzone et al. (2015, 2017) in turn focus on the societal dimensions of geothermal energy using the framework of responsible innovation and refer in particular to the importance of public engagement.

In light of this existing debate on the intersection between social acceptance and responsible innovation, the third publication of this dissertation aims to link the two concepts and research strands more closely using the concept of participation as a core element. To this end, an *empirical case study* will examine whether and how a conceptualisation of participation in the sense of responsible innovation could be useful for dealing with social acceptance and what implications this has for responsible

innovation. Based on an empirical case study on the social acceptance of green hydrogen in Germany, the article therefore aims to answer the following two research questions: (1) What role does participation play in the social acceptance of innovation (using the case of green hydrogen)? (2) What might an acceptance-oriented conceptualisation of participation mean for responsible innovation?

4. Publications

4.1 Aligning Innovation and Ethics: An Approach to Responsible Innovation Based on Preference Learning

4.1.1 Publication details

The first co-authored paper by Häußermann, J. J. and F. Schroth (2020) was published in *Philosophy of Management* (ISSN 2052-9597, volume 19, issue 3, pages 349-364). The postprint version of the paper is available via DOI: 10.1007/s40926-019-00120-1. Author contributions are: research design J.H., wrote the paper: J.H. & F.S.

4.1.2 Abstract

New technologies not only contribute greatly to society and the economy; they also involve fundamental societal shifts, challenging our values and ideas about ourselves and the world. With a view to aligning technological change and innovation with ethical values, the concept of responsible innovation advocates the inclusion of a variety of stakeholders, in particular from society. In shifting moral responsibility towards the producers of innovations, responsible innovation rejects the standard normative economic view that the ethical evaluation of innovations is a matter of individual consumers' market-based choices. However, in this article we argue that responsible innovation should not abandon all normative consideration of the individual outright, to which end we present an alternative normative economic approach based on preference learning. We show how this approach can provide an enhanced understanding of responsible innovation by clarifying the redistribution of moral responsibility and casting individuals in the normative role of co-innovators, rather than mere consumers. We argue that responsible innovation should enable individuals to form preferences and evaluate innovations, so as to align innovation with ethical demands. Finally, we show how our proposed approach can be put into practice in so-called laboratories in real-world contexts, using methods from the field of design. In short, drawing on normative economics, this article aims to establish a new understanding of responsible innovation that is conceptually sound and can form the basis for novel innovation practices.

4.1.3 Introduction

The tremendous impact of what is often described as digital transformation has sparked off new interest in and intense debate on ethics, technology and innovation. The new technological possibilities presented by Artificial Intelligence, Big Data, the Internet of Things and Industry 4.0 raise significant ethical challenges concerning such issues as human self-determination, the right to privacy and protection against discrimination. A fundamental challenge here is to determine which ethical principles should govern the responsible design of new technologies. Ultimately, this is underpinned by the question of what shape we want technological change to take: How do we as a society want to utilise new technologies? How can they best promote the common good? Which innovations do we want to shape our future, so as to enable individuals to live a good life? In the absence of an underlying ethical framework, technologies and innovations alone cannot answer questions concerning what future we want and what ultimately constitutes a good life (Mulgan 2016; Philbeck, Davis, and Larsen 2018). Therefore, the dynamics of digital transformation and the profound economic, social and ethical impact of digital innovations require new approaches that provide an ethical grounding for innovation and technological change.

In light of the above, this paper shows how the concept of responsible innovation can be supplemented with insights from normative economics in order to provide a tenable and comprehensive solution to the challenge of aligning innovation with ethics. Responsible innovation entails a redistribution of moral labour and a shift of responsibility towards the producers of innovations – in repudiation of the mainstream normative economic view that ethical evaluation of innovations takes place on the market, through individual consumers' choices with respect to new products or services. We argue, however, that in rejecting said mainstream normative economic view the dominant understanding of responsible innovation too readily dismisses the role and normative status of the individual. Instead, we suggest that responsible innovation could be conceived of in an even more participatory and inclusive manner, by incorporating the role of the individual – and hence society in general – via a different normative approach, based on the notion of preference learning. Our proposal thus corroborates the notion that individuals should not be regarded as mere consumers, and provides justification for instead seeing them as co-innovators.

We defend our proposed normative approach by showing that the aforementioned mainstream normative economic view, according to which innovations are subject to market-based evaluation as a result of consumers' choices, should be rejected for at

least two reasons. Our proposed preference learning-based normative economic approach, however, makes it possible to salvage the role of the individual and normative individualism, which can in turn be integrated into – and thereby enhance – the concept of responsible innovation.

Moreover, as we show towards the end of this paper, the approach we propose not only serves to refine the concept of responsible innovation, but has also enabled us to develop a novel methodological framework for putting said concept into practice, via so-called laboratories in real-world contexts.

In short, this article argues that responsible innovation should better integrate society at the level of the individual, on the basis of an alternative normative economic perspective centred on the notion of preference learning. This requires that individuals are given the opportunity and means to form preferences towards (potential) innovations – which can in turn serve as the basis for a novel practical approach to responsible innovation.

The paper is structured as follows. First, in Section 2, we briefly present the concept of responsible innovation and explain how it seeks to better align innovation with ethics. We then discuss conventional normative economics in Section 3. We start by raising two objections to the dominant normative perspective based on welfare economics, before presenting an alternative normative economic approach, which is based on preference learning and is applicable to innovation. In Section 4, we argue that this alternative normative approach should be integrated into the concept of responsible innovation, so as to provide a more comprehensive ethical framework for innovation. Finally, we present a practical methodological approach that enables individuals to develop and become of aware of their preferences towards (potential) innovations. We conclude by clarifying how we hope to have contributed to a better understanding of responsible innovation by means of our analysis and proposal, before suggesting some questions for further research.

4.1.4 Responsible Innovation as a Means of Aligning Innovation and Ethics

A new yet already fairly well-developed concept, responsible innovation seeks to better align innovation – broadly understood – with societal values and needs, in terms of both processes and outcomes (René Von Schomberg 2013). Combining elements of different disciplines and perspectives (Grunwald 2011; Häußermann and Heidingsfelder 2017; Lubberink et al. 2017), responsible innovation is concerned with whether innovations have a detrimental or positive impact on society, as well as clarifying what the "right" effects of given innovations may be (René Von Schomberg 2014). One particularly influential account of responsible innovation is provided by Stilgoe, Owen and Macnaghten (R. Owen, Macnaghten, and Stilgoe 2012; Richard Owen et al. 2013a; Stilgoe, Owen, and Macnaghten 2013). According to their framework, innovation processes that are designed in accordance with and governed by four central dimensions – anticipation, reflexivity, inclusion and deliberation, and responsiveness – produce outcomes that ultimately better address societal values and needs. Though this may currently be the dominant conceptualisation of responsible innovation, it is still a relatively young field, marked by the emergence of new and competing perspectives (Burget, Bardone, and Pedaste 2017; Lubberink et al. 2019; Ribeiro, Smith, and Millar 2017; Rip 2016; Timmermans and Blok 2018). Still, given the central aim of aligning innovation with societal values and needs, one feature that is common to all interpretations of the concept is the inclusion of different stakeholders, in particular the public, for the sake of more democratic innovation processes (Brand and Blok 2019; Taebi et al. 2014; Timmermans and Blok 2018). While more conventional forms of technology assessment address the social and ethical aspects of innovation only after a product or service has been developed and launched (Grunwald 2014), the concept of responsible innovation emphasises the inclusion of ethics within the innovation process itself (Brand and Blok 2019; Häußermann and Heidingsfelder 2017). To this end, a variety of approaches, ranging from methods based on design research (M. Heidingsfelder et al. 2015) to new collaboration strategies (Schroth and Häußermann 2018), have been developed, discussed and put to use. In seeking to democratise innovation by involving not only direct stakeholders, but also (lay) members of the public and other societal actors that may not have an immediate stake in a given innovation process (Richard Owen et al. 2013a), responsible innovation contributes to a new distribution of moral labour and, by extension, (co-)responsibility for innovations (Brand and Blok 2019; Rip 2014; Schlaile et al. 2018). In contrast to the conventional approach of leaving the ethical assessment of innovations to the market and individuals as consumers, responsible

innovation shifts responsibility towards the producers of the innovations in question, whether it be businesses, research organisations or government agencies. It thus treats the ethical assessment of innovations, i.e. asking whether they are good or bad and to what extent they can be considered societally desirable, as a participatory process that includes the general public, so as to shape technological change according to ethical and social principles – or as Brand and Blok (2019, 6) put it: "It is held that when diverse stakeholders and lay people are involved in the innovation process, innovators can learn about the social desirability of an innovation and become responsive to societal values and needs."

We believe that the concept of responsible innovation as described above is at base sound and presents an auspicious means of aligning innovation with ethics. However, while it is reasonable to shift the ethical assessment of innovations away from individuals in their roles as consumers and towards the innovators themselves, we argue that responsible innovation should not abandon the perspective of the individual altogether. In its rejection of the mainstream market-based normative perspective, the dominant understanding of responsible innovation overlooks alternative normative economic approaches to innovation that could help to incorporate the perspective of the individual (cf. Schlaile et al. 2018). In particular, we argue that responsible innovation should draw on a different kind of individual-based normative economic approach, namely one which focusses on how individuals form preferences towards innovations in the first place. Using the process of preference formation as normative starting point in turn provides the basis for novel practical implementations of responsible innovation.

In the next section, we develop precisely such an approach. We start by introducing the mainstream normative economic perspective on innovation, which is based on conventional welfare economics, and which seems to be assumed by the dominant understanding of responsible innovation. After discussing two objections to this view, we present an alternative approach that draws on the notion of preference learning and takes a more rounded view of the role of the individual, which can accordingly be integrated into an enhanced understanding of responsible innovation.

4.1.5 Innovation and Welfare Economics: Two Problems with the Conventional Ethical Assessment of Innovation

When trying to find a suitable normative criterion and justification for innovation, it makes sense to start by looking at how mainstream economics assesses particular effects, developments or states in terms of how well individuals, or society in general, is said to be faring – which in turn relies on a particular understanding of "welfare". Accordingly, in this section we briefly present the standard approach to ethical questions within economics, which draws on conventional welfare economics. We then put forward two reasons why this approach seems questionable, especially in the context of innovation. Our first argument is based on the particular epistemic quality of innovations. The second argument we present draws on the findings of behavioural economics, in light of which it seems unreasonable to assume that people have stable and coherent preferences. It is important to emphasise that an extensive theoretical discussion of normative economics falls outside the scope of this paper. Rather, our aim here is to problematise the standard approach in order to identify alternative perspectives on ethics and innovation, and so to arrive at a novel approach to responsible innovation.

4.1.5.1 Mainstream Economics and Normativity

Within mainstream economics, the ethical evaluation of policies, outcomes and institutions in terms of increases or decreases in welfare follows conventional welfare economics, which defines welfare as extent to which preferences are satisfied (Hausman, McPherson, and Satz 2017). In other words, this view equates individual well-being (i.e. welfare) with the satisfaction of personal preferences, all of which carry equal weight; interpersonal comparisons of preferences are considered neither possible nor relevant (Robbins 1932). Based on the premise that it is (morally) good if people are better off, conventional welfare economics therefore holds that, given that other normative demands are not at stake, it is good – and hence required – to satisfy people's preferences in order to make them better off. For welfare economics to constitute a robust framework of economic theory, it must therefore be able to provide a convincing account of what preferences are, and how they can be identified or assessed. This is where revealed preference theory, which underpins the conventional conceptualisation of preference within economics, comes in. The main assertion of this theory is that people's preferences can only be inferred based on their

actual observed choices; that is, whether or not they ultimately bought a new product, used a new service, or the like.

This depends on certain central assumptions about the nature of preferences. One such key assumption is that people's preferences are clearly defined, in the sense that they are in fact always able to decide between two alternatives, or are otherwise truly indifferent. Another important supposition is that people's preferences remain stable over time and do not change depending on different contexts (Hausman 1992).

It is important to note that this theory does not consider the specific reasons someone may have for preferring a given option – such as its usefulness or pleasurability – to be ethically relevant. Utility is defined purely as preference satisfaction and, accordingly, maximising utility is simply a matter of doing whatever one most prefers to do, irrespective why. Different options can thus be represented by an ordinal utility function, where the position of each given option is determined solely by its place in an agent's preference ranking (Hausman and McPherson 2009). In thus defining preferences as the ranking of alternatives that is implicit in people's choices, revealed preference theory can be understood as identifying utility maximisation with choice (Hausman 2012). Moreover, this theory can be understood as defining the conditions for rationality and rational choice (cf. Mas-Colell, Whinston, and Green 1995). On the premise that rational agents would strive to improve their welfare through maximisation of utility, and given that revealed preference theory equates utility with preference satisfaction, conventional welfare economics holds rational choices to be those which satisfy the agent's preferences. These preferences, in being the only relevant normative determiner of individual welfare, can be imputed by observing people's choices – which can in turn be predicted on the basis of their revealed preferences. Thus standard economic theory has ultimately come to identify preferences with choices, and to treat the terms "utility maximisation" and "choice" as synonymous (Gul and Pesendorfer 2008).

In short, conventional welfare economics sees the satisfaction of individual preferences as revealed in peoples' choices as sole normative criterion (Hausman, McPherson, and Satz 2017; Reiss 2013). Thereby, it relies on neoclassical assumptions of rationality, according to which people have stable and coherent preferences (Hausman 2012).

4.1.5.2 Two Objections

In the context of innovation, there are at least two objections that may be raised against the normative foundation described above.

The first pertains to an epistemic problem. Innovation involves novelty. As economics is concerned with individual choices (see above), innovations can be described as involving choices or actions that have no precedent (Witt 1996). Whether a product, service, or something else, an innovation is by definition something previously unknown, which has not been given prior consideration; thus, its value and effects cannot be anticipated in advance (Witt 1996). Taking this epistemic dimension of innovation seriously, it no longer seems reasonable to rely on a normative account which requires people to have stable and coherent preferences in the context of an innovation-based and thus constantly evolving economy. How should individuals have well-formed preferences towards something which is necessarily unknown and thus not assessable? It is difficult to explain why individuals should choose to adopt an innovation, say an entirely new product or service, for which they could not have developed a prior preference. Yet, as we have seen above, conventional welfare economics relies on the assumption that individuals have consistent, stable and well-formed preferences towards innovations, independent of changing contexts. This seems implausible, given that change is an integral part of the concept of innovation. Given that innovations entail novelty, change and perhaps even disruption, it is unreasonable to expect individuals to have specific preferences with respect to, for instance, a new product that they have yet to get to know. Moreover, the formation or adjustment of preferences with respect to an innovation will more often than not impact an individual's existing set of preferences (Binder and Witt 2011), which makes it problematic to refer to existing preferences in the normative evaluation of (new) innovations. As Binder and Witt (2011, 6) argue, individual preferences may be contingent on past innovations, which would mean that preferences are systematically affected by innovations: "The major difficulty is that preferences tend to become endogenous to the process of innovative change in the economy: the preferences by which individual well-being is assessed are shaped through the very processes whose welfare effect they are supposed to evaluate." This process is illustrated in Elster's (1983) account of how the Industrial Revolution may have led to less welfare as a result of an increase in opportunity sets, since people were no longer able to satisfy their preferences to the same extent: "We were happier before we got these fancy new things, but now we would be miserable without them" (Elster 1983, 135). Since, as we have just shown, preferences are systematically affected by innovations, it is untenable

to base the normative assessment of innovations on the satisfaction of preferences. How should the welfare effects of innovations be assessed on the basis of preferences, if these preferences are themselves shaped through innovations?

In short, innovation that requires novelty – and thus genuine uncertainty – is incompatible with conventional welfare economics, which treats the satisfaction of pre-existing and perfectly consistent preferences (as revealed through people's choices) as normative criterion (Schubert 2012, 2015a).

Our second argument against the normative approach to innovation within conventional welfare economics centres on a behavioural problem. Going as far back as the 1970s, the findings of behavioural economics have challenged, or even refuted, the validity of neoclassical axioms of rationality as basis for an adequate positive description of human behaviour (Kahneman, Slovic, and Tversky 1982; Tversky and Kahneman 1974). Since, as behavioural economics shows, people (often) lack stable and coherent preferences, their behaviour cannot provide reliable evidence of their welfare – in contrast to what revealed preference theory assumes. Due to cognitive limitations, people often fail to satisfy their preferences and make inferior decisions regarding their own welfare. That has led some to argue that people's inconsistent choices may be treated as mistakes (Sunstein and Thaler 2003; Thaler and Sunstein 2003). This line of reasoning, however, leads one to notions of "preference purification" and "informed" or "true" preferences (Hausman 2012; Infante, Lecouteux, and Sugden 2016b, 2016a; Lecouteux 2015; Sugden 2015b), which are subject to serious objections. These notions view human decision-making in terms of psychological mechanisms that interfere with rational choice. However, though it explains deviations from latent reasoning, this approach provides no psychological explanation of the process of latent reasoning itself – which is problematic, to say the least (cf. Kahneman 1996). Thus, there is, or at least seems to be, no good reason to believe that informed, true or latent preferences exist (Whitman and Rizzo 2015). The fact that innovation occurs in a highly dynamic context that involves change and uncertainty makes the assumption of the existence of stable and "true" preferences particularly questionable.

In short, attempting to adhere to conventional welfare economics within a behavioural context reveals a conflict between normative and behavioural economics (McQuillin and Sugden 2012). The conventional normative view of preference satisfaction relies on the assumption of the existence of some sort of "true" preferences, which people would have if they were free from behavioural bias. However, as there is no evidence

that such preferences exist, this view is unfounded (Sugden 2015a). In other words, it is not reasonable to assume that people have stable and coherent preferences towards innovations, on the basis of which those innovations may be evaluated.

We have argued that the findings of behavioural economics carry major implications for normative economics – in particular conventional welfare economics, which treats the satisfaction of preferences as normative criterion. Although this argument can be expanded and developed into a general criticism of conventional welfare economics, it is used here merely to indicate the crucial shortcomings of a normative approach based on the satisfaction of preferences. This is particularly relevant to current discourse on innovation, since the assumption of the existence of "true" preferences has significantly influenced how innovation is conceived of, presented and ultimately implemented. This kind of normative understanding affects how innovation economies are shaped and what innovators implicitly think they should do – yet as we have shown, in the context of constant economic and technological change, welfare cannot be defined in terms of existing preferences. Therefore, there is a need for new normative criteria, which take into account the very nature of innovation.

In this section, we first introduced the standard economic approach to assessing innovations, which is based on welfare economics and uses the satisfaction of individual preferences as normative criterion. We then presented two reasons why said normative approach is problematic. While the first criticism we presented focusses on the particular epistemic nature of innovation, our second objection draws on key insights from behavioural economics, and their implications for normative assessment on the basis of preference satisfaction. In view of the fact that the normative dimensions of innovation currently form the subject of heated debate, and given that appropriate ethical criteria are urgently needed, the absence of a tenable normative approach within economics is striking (cf. Binder and Witt 2011; Schubert 2012).

4.1.5.3 Preference Learning as Normative Criterion

In order to avoid both of the problems with conventional welfare economics discussed above, we propose that the focus be shifted from single preferences to the dynamic process that gives rise to those preferences in the first place (Dold and Schubert 2018) as the basis for normative evaluation. Rather than assuming that people already have stable and coherent ethical preferences with respect to innovations, we maintain that it is important that they are given the space and opportunity to develop ethical

preferences, so that innovations can be aligned with their needs and values.[9] Hence, we follow Schubert (2012, 2015a, 2015b) and Witt (2001) in arguing that the relevant normative criterion should be the possibility for individuals to "learn"; that is, to experiment and acquire preferences as part of their personal development. In other words, while it is beyond the scope of this paper to develop a new theoretical model of normative economics, we argue that preference learning provides the grounds for a plausible normative approach that can be integrated into the concept of responsible innovation. Accordingly, rather than the satisfaction of people's imputed preferences, what matters is whether individuals can learn, interact with innovations, experiment with new preferences and thus "creatively construct their own identity" (Dold and Schubert 2018, 233) on an ongoing basis. This corresponds to Schubert's (2012, 2015b) call for "effective preference learning" as a new conceptualisation of welfare or well-being, to be used as normative measure in an innovation-based and evolving economy. This procedural concept of welfare is to be understood as the capacity "to engage in the ongoing learning of instrumentally effective preferences" (Schubert 2012, 585). Thereby, we also avoid some kind of substantive assumption about the normative criterion evaluating innovations. Underlying this procedural understanding of welfare as a continuous preference learning process is the idea that agents adapt to change through learning. Witt (2001) and Earl and Potts (2004) maintain that although in the face of complexity, ignorance and uncertainty, human cognitive capacity may never allow individuals to be aware of a complete set of preferences, they can nonetheless engage in the continuous construction of their respective dynamic preference sets. As Schubert (2012) argues, "individuals adapt to changing circumstances by trying out and acquiring new preferences." In this way, they act as "experimental consumers" with the objective of continuously exploring both their needs and sources of utility to satisfy said needs. Assuming that individuals construct their identity by reflecting upon their evolving preferences, Dold and Schubert (2018) claim that individuals care about being able to form their preferences in a self-determined way (Deci and Ryan 2000), and thus argue that individuals should be conceptualised as "loci of learning" (2018, 234). Accordingly, the account of

[9] By "ethical preferences", we mean individual preferences with respect to the ethical and social implications of a given innovation. In other words, we do *not* use the term to draw a distinction between ethical and unethical preferences, but rather to refer to an individual's stance, attitude or personal values in relation to a particular innovation. Any given innovation can thus be understood to have an ethical dimension, which is made visible and becomes a deciding factor as a result of people's ethical preferences with respect to it.

preference learning that we are propounding focusses on how individuals actually develop their preferences or, in other words, how they "learn".

But how does the concept of preference learning avoid the two objections raised above? First, the inherent novelty of innovation no longer poses a problem, since the shift in focus to the process of preference formation means that single preferences are not required to be stable and coherent. On the contrary, it is the dynamic process of preference formation that enables individuals to engage with the unknown. Therefore, requiring that the process of innovation be aligned with that of preference formation is a tenable normative criterion that takes into account the novelty and uncertainty inherent in innovation. Secondly, our preference learning-based approach avoids the problem of assuming the existence of "true" or latent preferences by changing the role and normative status of single preferences to outcomes of an ongoing process of learning (cf. Dold and Schubert 2018, 236). In other words, by shifting the focus to the process of preference formation – i.e. the way individuals reflect on, balance and negotiate their eventual preferences – we are able to overcome the dualistic model of a consistent, reflective and "true" self on the one hand, and a fast and intuitive self on the other (cf. Kahneman 2011).

Finally, the notion of preference learning as described above can be integrated into responsible innovation and provide a novel normative economic substantiation for the concept. Focussing on the process of preference learning not only takes the participatory dimension of responsible innovation seriously, but also enhances our understanding of the concept by providing a new perspective on the role of the individual.

We furthermore argue that conceiving of responsible innovation in terms of preference learning provides a solid foundation for the development of practical measures that enable individuals to construct and reflect on preferences. Or, as Dold and Schubert (2018, 237) maintain, policies to further preference learning "should focus on the task of modifying those contextual elements that foster individuals' means of preference learning, viz., *educational and cultural institutions* that provoke individuals' reflection of their evolving preferences and *access to markets* that allow individuals to test different preferences on a continuous basis." Accordingly, in the next and final section of this paper we present a methodological approach based on our proposed understanding or responsible innovation, which aims to foster preference learning.

4.1.6 Laboratories in Real-World Contexts: A Practical Approach to Responsible Innovation Based on Preference Learning

As shown above, the concept of responsible innovation addresses the challenge of aligning innovation with societal values and needs through a democratisation of innovation via the inclusion of different and diverse stakeholders – in particular societal actors. This leads to a redistribution of responsibility and a new division of moral labour. However, as we have shown, the heretofore dominant understanding of responsible innovation has largely neglected the role and normative status of the individual. We therefore propose a normative economic clarification and justification of the role and inclusion of individuals that draws on the notion of preference learning – and thus builds on and increases the participatory aspect of responsible innovation. While this is in keeping with certain central aspects of the notion of responsible innovation, namely the rejection of the conventional idea of market-based normative judgements on the part of individuals as mere consumers, and a shift in responsibility towards the innovators themselves, our proposed approach nonetheless assigns a crucial role to the individual – albeit not as a consumer, but rather as a person, member of society and co-innovator. In short, conceiving of responsible innovation on the basis of preference learning entails understanding participation as more than the mere involvement of societal stakeholders. It is crucial that individuals are given the means to "learn"; that is, to form (ethical) preferences towards innovations[10] and reflect upon their evolving preferences with respect to, for instance, emerging technologies. This kind of learning requires an appropriate institutional and methodological framework which, as we demonstrate below, can be provided by so-called laboratories in real-world contexts (LRWs) (cf. Schäpke et al. 2018).

As part of an experimental turn in the social sciences (Overdest, Bleicher, and Gross 2010), LRWs have gained in prominence over the last two decades. Examples of LRWs include living labs, real-world laboratories, social labs and other similar

[10] Note that by "ethical preferences", we mean individuals' preferences with respect to the ethical and social dimensions of innovations (see footnote 1). Responsible innovation thus focusses on the ethical dimension of a given new technology when it seeks to evaluate whether or not, to what extent or in what form it might be "preferred". Although our notion of ethical preferences could be contextualised within current debates about hierarchical preferences (cf. Lazar and Klein 2018; Schnellenbach 2019), doing so falls beyond the scope of this paper, as it would shift the focus to a theoretical discussion of behavioural economics. Nevertheless, the relationship between hierarchical or reflective preferences and the notion of ethical reflection in responsible innovation may be worth investigating in future research.

approaches, all of which are marked by the collaboration of scientific and societal actors, an embeddedness in real-world contexts, and experimentation (Schäpke et al. 2018). They aim to facilitate the collaborative development of solutions through new forms of cooperation, dialogue and reflection in novel setups (Hassan 2014; Kieboom 2014).

LRWs qualify as an environment to enable preference learning due to three reasons: first, they enable the integration of ethical concerns in technology trajectories at the stage of the research process. In such an environment, individuals may become co-innovators as LRWs provide an opportunity to incrementally adjust science and innovation as they actually occur (Richard Owen et al. 2013a). Second, in contrast to the closed setting of secluded laboratories (Callon 2009), LRWs move research from in-vitro to in-vivo settings (Schaer 2017). Thus, interactions and dynamics of the real world can be taken into account and research opens up to groups outside the laboratory (Schroth, Glatte, and Kaiser 2020). Regarding the integration and acknowledgement of ethical issues in R&D processes, this is advantageous compared to laboratories, as ethical issues "cannot be tackled within the laboratory [and] should be put on the agenda of groups outside the laboratory" (Boenink 2013). Third, LRWs are marked by multi-stakeholder involvement and experimental learning, reflecting some of the core principles of responsible innovation, i.e. reflection, deliberation and inclusion. They might therefore seem like good candidates for the practical implementation of an understanding of responsible innovation that incorporates preference learning.

Upon taking a closer look at these concepts, however, it becomes apparent that they fall short in terms of preference learning. The objective of living labs, for example, is to better understand the complexity of a given real-world environment and users' interaction with technologies, whereas social labs are used to initiate the development of social innovations and to investigate people's experiences with these innovations in real-world settings (Schroth and Schraudner 2019). Both of these approaches lead to a better understanding of the complexity of the real world than would be afforded by the insulated setting of a traditional laboratory. However, in both cases the "learning" that occurs is focussed on enhancing the functionality of the innovation at hand, with ethical considerations and preference learning playing only a marginal role. Furthermore, in both instances, it is primarily the researchers who "learn", in that they gain new insights and information. As we have argued above, however, the full potential of responsible innovation can only be realised if the users of future

technologies, products and services are also given the means to "learn", i.e. to develop and reflect on their preferences with respect to these innovations.

We thus suggest the following methodological refinements, which should serve to better align the core principles of LRWs with responsible innovation, and to create the conditions required for preference learning.

To make preference learning – and hence responsible research – possible, we suggest that LRWs create the opportunity for participants to come together physically in order to articulate their needs, formulate ideas, and reflect on their preferences. By employing a set of interdisciplinary methods from the field of design and from the social sciences, participants can be given the means to make possible and necessarily unknown future scenarios visible and tangible (for a more detailed discussion of these methods, see for example: M. Heidingsfelder et al. 2015; M. L. Heidingsfelder, Schütz, and Kaiser 2016; Kaiser et al. 2019). In particular, experimentation with and co-design of so-called speculative prototypes of new technologies, products and services can provide a basis for interpersonal exchange and social debate on the normative implications and potential ethical trade-offs entailed by specific innovations. Unlike in approaches such as living labs, these prototypes are used not to improve the functionality of, for instance, newly-developed products, but instead to enable individuals to articulate, negotiate and reflect on their needs and ethical stance with respect to potential innovations – and thus to "learn" about their own ethical preferences.

This approach was used in an LRW focussed on innovation for rural development, where participants encountered speculative prototypes of future technologies on a walking tour through their hometown. At the existing bus station, for instance, they were confronted with a prototypical possible future mobility station, where they were encouraged to think about the potential applications of future mobility solutions – and of advantages and disadvantages these solutions might hold for them personally. While discussing these perceived pros and cons, the participants gradually became aware of their respective individual preferences with regard to possible future mobility solutions.

Alternatively, speculative prototypes can take the form of desirable future scenarios. Combining methods and tools from participatory design with established foresight practices, the development of these scenarios encourages individuals to think about future technologies and innovation without being bound by practical constraints – and thus to explicate preferences they would not be able to formulate when thinking in

terms of existing limitations. This, too, serves to create open discourse and enable individuals to identify their own preferences (Kaiser et al. 2019).

In the aforementioned LRW on rural development, for instance, participants worked in teams to design desirable future scenarios for rural areas. In doing so, they debated and grappled with the meanings, implications and crucial ethical aspects of potential future innovations. Thus, for instance, participants formulated and articulated their individual preferences with regard to autonomous vehicles and sharing platforms. They then presented those preferences to other workshop participants.

As these examples show, our proposed practical approach brings multiple stakeholders together physically to experiment with future and speculative technologies and innovations in a real-world context, which enables them to form and become aware of their own individual preferences. In other words, the type of LRW we propose combines the two central principles of social and living labs, namely multi-stakeholder engagement and experimental learning, in a manner that makes the formation of individual preferences possible.

Moreover, helping individuals develop ethical preferences with respect to the (future) development of technologies and innovations creates the possibility of shaping the future on the basis of said ethical preferences. Our proposed practical approach to responsible innovation could thus be considered as an example of the kind of ethics foresight Floridi calls for, i.e. a soft and translational ethical approach to new digital technologies (Luciano Floridi 2014, 2018; Taddeo and Floridi 2018) – which is very similar to what we propose in this paper. In other words, the practical approach to responsible innovation we have presented here makes it possible to align innovation and technological change with societal needs and values and, ultimately, for us to work towards the kind of future we want.

4.1.7 Conclusion

This paper introduced the concept of responsible innovation as a means of aligning innovation with ethics. Noting that in seeking to redistribute moral responsibility for innovation, responsible innovation rejects the view that innovations should be assessed through individuals' market-based choices, we argued that the role of the individual should nonetheless not be dismissed altogether. Drawing on the notion of preference learning, we proposed a new normative economic approach, on the basis of which responsible innovation could incorporate a more rounded perspective on the role of the individual and achieve a greater level of participation. Moreover, as we hope to have demonstrated, our approach is compatible with actual innovation practices.

In line with Blok (2018b) and Blok and Lemmens (2015) in this paper we aimed to clarify some of the important conceptual assumptions underlying innovation. However, while we did try to integrate findings from economics into our analysis and of and proposed modifications to responsible innovation, we call for more attempts to include perspectives from economics, business, and innovation management in future research on the topic. For example, further studies might ask how responsible innovation could be integrated into conventional innovation management, or what advantages it offers in a competitive marketplace. Finally, following Timmermans and Blok (2018), we explored possible practical implementations of our proposed understanding of responsible innovation. Nevertheless, we see a need for further research that translates responsible innovation's conceptual principles into actual innovation processes, and for corresponding empirical analyses of said processes and their outcomes.

To conclude, with our proposed expanded understanding of responsible innovation, and the corresponding practical framework for integrating ethical preferences into innovation processes by means of LRWs, we have taken the first steps towards a better alignment of innovation with ethics. We believe that our approach to responsible innovation presents a promising solution to the challenge of integrating ethical preferences into innovation processes, so as to enable society itself to set the course for digital transformation and the use of new technologies. Thus innovation becomes not just the furtherance of (inevitable) technological and economic progress, but rather a process of actively shaping the future in accordance with societal norms and values.

Acknowledgements

This paper was presented at the 13th Philosophy of Management conference, which took place in June 2018 in London. The authors wish to thank the organisers and participants of the conference, in particular the track chairs Vincent Blok and Job Timmermans, for valuable feedback.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

4.2 Community-in-the-loop: towards pluralistic value creation in AI, or—why AI needs business ethics

4.2.1 Publication details

The second co-authored paper by Häußermann, J. J. and C. Lütge (2021) was published in *AI and Ethics* (ISSN 2730-5961) as part of the topical collection "Governance and AI: Challenges for a sustainable digital ecosystem". The postprint version of the paper is available via DOI: 10.1007/s43681-021-00047-2. Author contributions are: research design J.H., wrote the paper: J.H. & C.L.

4.2.2 Abstract

Today, due to growing computing power and the increasing availability of high-quality datasets, artificial intelligence (AI) technologies are entering many areas of our everyday life. Thereby, however, significant ethical concerns arise, including issues of fairness, privacy and human autonomy. By aggregating current concerns and criticisms, we identify five crucial shortcomings of the current debate on the ethics of AI. On the threshold of a third wave of AI ethics, we find that the field eventually fails to take sufficient account of the business context and deep societal value conflicts the use of AI systems may evoke. For even a perfectly fair AI system, regardless of its feasibility, may be ethically problematic, a too narrow focus on the ethical implications of technical systems alone seems insufficient. Therefore, we introduce a business ethics perspective based on the normative theory of contractualism and conceptualise ethical implications as conflicts between values of diverse stakeholders. We argue that such value conflicts can be resolved by an account of deliberative order ethics holding that stakeholders of an economic community deliberate the costs and benefits and agree on rules for acceptable trade-offs when AI systems are employed. This allows AI ethics to consider business practices, to recognise the role of firms, and ethical AI not being at risk to provide a competitive disadvantage or in conflict with the current functioning of economic markets. By introducing deliberative order ethics, we thus seek to do justice to the fundamental normative and political dimensions at the core of AI ethics.

4.2.3 Introduction

Today, due to growing computing power and the increasing availability of comprehensive, high-quality datasets, so-called artificial intelligence (AI) technologies are increasingly being used in almost all sectors and are thus entering many areas of our everyday life (Perrault et al. 2019; Benaich and Benaich 2019). Yet the use of AI-based algorithmic systems raises ethical questions, calls societal beliefs into question and challenges many fundamental values (Luciano Floridi et al. 2018; Dignum 2018). This concerns, for example, questions of discrimination and fairness, privacy and human autonomy in semi-automated decision-making, risks of individual and social surveillance or threats to democracy through dynamic misinformation in social media and to human life through autonomous weapon systems or drones (B. D. Mittelstadt et al. 2016; Tsamados et al. 2020). Addressing the complex social, ecological and ethical consequences the development and use of AI systems might have, the emerging field of AI ethics seeks to establish normative approaches both on a theoretical as well as practical level which mitigate adverse effects and enhance the advantages of AI for the benefit of society.

Bringing together several different concerns about its evolution, we identify five crucial shortcomings of the first two waves of AI ethics. Based on this analysis, we introduce a business ethics perspective of deliberative order ethics claiming that at the core the use of AI systems may lead to fundamental value conflicts which to resolve AI ethics needs to be adequately equipped. In short, we argue that by too narrow a focus on technical systems, current AI ethics tends to ignore the context of using AI, namely their integration into business practices and economic markets. The question then becomes how AI ethics could include a broader normative perspective which acknowledges the wider societal embeddedness of AI innovation. In response to this question, we advocate complementing AI ethics with a normative theory of business ethics that makes it both theoretically more solid and practically better applicable given the conditions under which AI innovation is (mostly) carried out today. Specifically, we present a contractualist approach of deliberative order ethics which stipulates that value conflicts triggered by the use of AI systems should be resolved by the stakeholders of an economic community deliberating and agreeing on mutually beneficial rules for balancing benefits and costs and acceptable trade-offs between diverse values. In this way, the ambition is to make the use of AI a matter of pluralistic value creation. Thus, acknowledging the political dimension of AI ethics, our approach of deliberative order ethics helps to address the fundamental normative questions raised by the use of AI in society (B. Mittelstadt 2019; Greene, Hoffmann,

and Stark 2019; Whittlestone et al. 2019; Green 2020; Rahwan 2018; Binns 2018; Wong 2020).

This article proceeds as follows: in the following section 2, we first outline the evolution of the first two waves of AI ethics before we then aggregate five crucial shortcomings at the threshold of an emerging third wave. The next section 3 proceeds by introducing the normative theory of order ethics and refining it in contrast to integrative social contracts theory (ISCT) as most proliferated theory of contractualist business ethics. Building on this, we then develop the concept of deliberative order ethics and discuss our approach in light of similar existing reasonings of the AI ethics debate (4). We then examine whether or not and to which extent our proposal may successfully address the five shortcomings identified in section two (5). We conclude by summarising our reasoning and highlighting both its purpose and relevance as well as its limitations (6).

4.2.4 Toward a third wave of AI ethics

4.2.4.1 From principles to practice

Although AI today is a highly interdisciplinary field, it can be described as a subfield of computer science which includes a range of technologies to create algorithmic systems that aim to reproduce human capabilities of intelligence (Mccarthy et al. 2006). Already established as a field of academic research since the 1950s, recent increases in computing power and the growing availability of large datasets allowed disillusions of the 1970s and late 1980s known as AI winters to be overcome. Today, it is particularly methods of machine learning and so-called neural networks that enable self-learning systems to be developed which, trained with the corresponding data, can ultimately perform even relatively complex tasks (Russell and Norvig 2016). Based on different techniques of learning such as supervised, unsupervised, reinforcement or deep learning, AI thus allows the creation of algorithmic systems that assist humans by their ability to perform tasks in a highly adaptive and (semi-)autonomous manner. AI systems are already widely used in almost all sectors of society, from manufacturing, agriculture, trade, finance and medicine to government and public administration. Applications range from digital assistants such as chatbots, language translation tools, recommender systems of varying complexity in the consumer sector or professional contexts, to applications for autonomous driving or complex robotic systems and face recognition technologies. However, the enormous potential and the broad range of possible applications do not only promise economic and business value. Often there are far-reaching social consequences for individuals and society as well as the environment. Ethical issues in the development and use of AI systems are raised, for example, with regard to the protection of individual rights, autonomy and privacy, risks of biases and discrimination based on characteristics such as skin colour, race or gender, the lack of accountability of AI-supported decisions, or risks of undesirable individual or social surveillance. Mittelstadt et al. (2016) and recently Tsamados et al. (2020) describe six types of ethical concerns. In addition to traceability, these include epistemic concerns about inconclusive, inscrutable or misguided evidence on the one hand and normative aspects such as unfair outcomes and transformative effects on the other. One well-known example is the case of a recruiting tool developed by Amazon which was designed to identify the most suitable candidates among the applicants based on data on previous career paths within the company. However, as the system revealed to discriminate heavily against women and systematically favoured male applicants, Amazon had to withdraw it completely. Another high-profile case is provided by COMPAS, a system designed to help courts

assess the risk of recidivism of defendants. Despite a high overall accuracy, however, it turned out that the probability of being wrongly assigned a high risk of recidivism was twice as high for a black offender than for a white offender ('false positive'), while white offenders were twice as likely to be wrongly assigned a low risk ('false negative') (Larson et al. 2016).

In view of the increasing use of AI and its vast influence on individuals and society, the debate about its ethical implications has attracted growing attention from the public, businesses, the academic community, and politics. In order to harness the benefits of AI while at the same time taking appropriate account of the ethical risks involved, a number of different actors from science ('The Montreal Declaration' 2017; Future of Life Institute 2018), politics (High-Level Expert Group on Artificial Intelligence (HLEG) 2019; AI4People 2018; OECD 2019; UK House of Lords 2017; Datenethikkomission 2019), industry (Deutsche Telekom 2018; Microsoft 2018; Google 2018), as well as professional associations (IEEE 2017) and civil society (Partnership on AI 2016) have developed principles and guidelines to enable the ethical and responsible use of AI. Although their focus varies in detail, cross-cutting issues and trends can be identified. Jobin et al. (2019) summarise a total of eleven, Hagendorff (2020) six and Floridi et al. (2018) five overarching principles. Using different review methods, Jobin et al. (2019) and Hagendorff (2020) highlight the principles of transparency, fairness or accountability. The principle of transparency, for example, aims primarily at disclosing the functioning of AI systems to make results explainable and interpretable. In this way, damage can be averted, (legal) justifiability verified and trust strengthened (Jobin, Ienca, and Vayena 2019). The principle of fairness seeks to prevent undesirable bias and resulting forms of discrimination in order to ensure diversity and equality. Accountability aims to ensure that decisions are justified in a comprehensible manner and that the distribution of responsibility is clarified in advance. From a more integrative perspective, the different principles and guidelines have been summarised with regard to established principles of bioethics of beneficence, non-maleficence, autonomy, justice, and explicability (Luciano Floridi and Cowls 2019; Luciano Floridi et al. 2018). While beneficence is to ensure that the use of AI promotes overall wellbeing and is consistent with sustainability and the common good (Luciano Floridi et al. 2020), the principle of non-maleficence aims to prevent potential damage caused by the use of AI (Luciano Floridi et al. 2018). In view of (semi-)autonomous systems, the principle of autonomy stipulates that people should always retain the last decision-making power or "the power to decide which decisions to take". Justice encompasses the effects that AI systems have on societies in

terms of unfair discrimination, but also on social cohesion and solidarity, and aims to ensure that the costs and benefits of the use of AI systems are fairly distributed within society (Luciano Floridi et al. 2018; Tasioulas 2019). Finally, the principle of explicability, which is the only one specifically for the context of AI, shall ensure that users and those affected by an AI system are able to understand and comprehend its results and that the distribution of responsibility is clear. We summarise the quest for principles and guidelines as the first wave of AI ethics. In view of the increasing use and impact of AI systems on individuals and society, it reflects the need to develop and use AI systems in line with a set of ethical values.

Even though the transition is certainly fluid, we take approaches tackling a concrete implementation of ethical AI as a second wave of AI ethics. One influential case is the ACM Conference on Fairness, Accountability and Transparency (ACM FAccT), formerly ACM FAT*, which evolved into an active community concerned with the ethical design of AI in close connection to relevant technical issues. In particular questions of explainable AI (Gilpin et al. 2019; T. Miller 2019; Madumal et al. 2018; Arrieta et al. 2019; Páez 2019; Gunning 2019) or issues of fairness (Seng and Lee 2021; Hellman 2020; Holstein et al. 2019; Barocas, Hardt, and Narayanan 2019; Kleinberg 2018; Chouldechova and Roth 2018; Pessach and Shmueli 2020; Corbett-Davies and Goel 2018) have emerged as productive fields of research. But also more governance-oriented approaches to the practical implementation of ethical AI play an important part, for example with regard to a professional code of conduct for developers (Boddington 2017; Bertelsmann Stiftung 2018), a more direct involvement of ethicists in the development of AI systems (Bonnemains, Saurel, and Tessier 2018; McLennan et al. 2020), or in terms of checklists (Gebru et al. 2020; Madaio et al. 2020), adapted internal structures (Rakova et al. 2020), suitable impact assessment frameworks (Schiff et al. 2020) and auditing processes (Raji et al. 2020) or a value-based AI label (AI Ethics Impact Group 2020). Finally, perspectives from the law concern the ethical design of AI at the interface with regulatory issues (Wachter, Mittelstadt, and Russell 2018; Calo 2018; Wachter and Mittelstadt 2019; Larsson 2020; Coeckelbergh 2019). In summary, Morley et al. (2020) provide a comprehensive overview of a variety of approaches and tools for the integration of ethical aspects in the development of AI systems. They develop a typology which relates the different approaches to implementing the five overarching principles according to Floridi et al. (2018), and assigns them to seven phases of an algorithmic development process. Overall, we conclude that a first wave of AI ethics, in view of the impact on individuals and societies, has put forward appropriate ethical principles

to guide the development and use of AI systems. The second wave builds on this and looks into how principles can be implemented and how guidelines can be put into practice. Although the difficulty of operationalisation and practical implementation is often emphasised (Jobin, Ienca, and Vayena 2019), the variety of approaches presented indicate that there are nevertheless a number of promising efforts in progress.

Based on the first and second wave in AI ethics described above, we argue that there are indications of a third wave, the upshot of which is not yet clear. Based on critical analyses of its evolution, we identify five key shortcomings of current AI ethics which we discuss in the next section.

4.2.4.2 Five shortcomings of current AI ethics

Following the quest for appropriate ethical principles and initial considerations on their practical implementation, a number of concerns have been voiced about the ensuing trends in AI ethics. In the following we will consolidate different concerns in order to delineate the current status of AI ethics. We argue that current critique can be summarised under five key shortcomings. First, AI ethics neglects the importance of business practices, without which, however, the ethical assessment of the use of AI systems is based on an incomplete picture (Greene, Hoffmann, and Stark 2019; B. Mittelstadt 2019; Veale 2020; Green 2020; Hagendorff 2020). Secondly, AI ethics is characterised by a form of technical solutionism which not only narrows the view of problems but also of options for action (Morley et al. 2020; Metcalf, Moss, and Boyd 2020; Greene, Hoffmann, and Stark 2019; Green 2020; Hagendorff 2020; B. Mittelstadt 2019). Closely related to this we find, thirdly, a focus on individuals, both in terms of the effects of AI systems and the responsible actors (Hagendorff 2020; Morley et al. 2020; Green 2020; B. Mittelstadt 2019; Greene, Hoffmann, and Stark 2019). Fourthly, the principle-focused approach of AI ethics faces problems in its practical implementation, on the one hand with regard to the necessary operationalisation of general principles, and on the other hand in terms of accountability and guaranteeing the intended effects (B. Mittelstadt 2019; Green 2020; Jobin, Ienca, and Vayena 2019). And finally, the unclear relationship between AI ethics and the legal regulation of AI is criticised, which, among other things, leads to AI ethics being misused by powerful corporations to prevent or at least delay legal action (B. Mittelstadt 2019; Rességuier and Rodrigues 2020; Jobin, Ienca, and Vayena

2019; Benkler 2019). In the following, we discuss these five shortcomings and their relevance to the field of AI ethics.

1. AI ethics neglects the business context of developing and employing AI systems in society

Although seemingly trivial, it is worth noting that it is mostly firms that commercialise AI systems and introduce them to markets at the end of an innovation cycle. Against this background, it seems reasonable to assume that an ethical assessment should take into account the business context of AI systems. In fact, even if an AI system is completely ethically designed on a technical level – if this is possible at all and whatever that may mean with regard to say fairness, privacy or safety in particular (Kleinberg 2018; Binns 2020; Wong 2020) – major ethical questions may arise. Think, for example, of the risks of dual use or the cases in which employees from Google or Microsoft have voiced public protest against the potential use of some of their companies' products for immigration and law enforcement agencies, military purposes or foreign governments (K. Martin, Shilton, and Smith 2019; Veale 2020; Whittaker, M., Crawford, K., Dobbe 2018). Or take the already widespread and various use of AI systems in recruiting, which raise questions about whether decisions about the future of people based on (psychological) profiling are legitimate and desirable. When and how are (semi-)automated decision-making processes about people's career prospects and opportunities for personal development societally desirable? Or, as Tasioulas (2019), p. 65 puts it: "Consider, for example, the plight of long-term unemployed people whose job applications are routinely rejected by the automated systems that now dominate workforce recruitment. After months or even years of applying unsuccessfully for jobs, those individuals may never once have their application read and evaluated by a fellow human. Even if we assume that the relevant algorithm meets a good standard of functionality, i.e. it is just as effective, efficient and compliant with norms of appropriateness as the average human recruiter, the fact that it is a non-human mode of decision-making is worrisome. It is hard to pin down the worry very precisely, but the thought is roughly that the job seeker is subjected to a cold, alienating, and ultimately potentially disrespectful process because his application never comes to the attention of a fellow human being. So much is suggested in this extract from a recent *Guardian* article: "It's a bit dehumanising, never being able to get through to an employer," says Robert, a plumber in his forties who uses job boards and recruiters to find temporary work. Harry, 24, has been searching for a job for four months. In retail, where he is looking, "just about every job" has some sort of test or game, anything from personality to maths, to screen out applicants. He

completes four or five tests a week as jobs are posted. The rejections are often instant, although some service providers offer time-delay rejection emails, presumably to maintain the illusion that a person had spent time judging an application that had already failed an automated screen.' (Buranyi 2018)". Hence, beyond issues of fairness or privacy questions arise as to whether its use may lead to ethically questionable business models, such as e.g. attention hacking (Greene, Hoffmann, and Stark 2019), whether the use of certain infrastructures such as cloud services directly or indirectly promotes competition-distorting monopoly structures (cf. Haucap 2018) or whether power balances relating to existing infrastructures are shifted through the use of AI (Veale 2020).

The problem of too narrow a view can also be substantiated from a more technical perspective. Using AI methods such as machine learning, so-called optimisation methods are often applied, which can calculate different models and optimisation functions on the basis of training data and defined optimisation goals, since an analytical solution to the problem is not possible (Kulynych et al. 2018; Gürses, Overdorf, and Balsa 2019; Datenethikkomission 2019). The use of optimisation technologies as a central element of AI systems shows two things. On the one hand, it illustrates that focussing on the individual protection of, say, privacy on a technical level does not allow the dynamic effects to be controlled in terms of profiling or manipulation of groups or societies (Veale 2020). Even if this means that companies are less interested in qualitative insights into individual data but only need the data for the statistical, probably even decrypted optimisation of services (Veale 2020), this shows all the more that AI ethics' focus on the technical improvement of the system itself does not grasp the full picture. Instead it is crucial to include questions about the acceptability of consequences, potential side-effects and the legitimacy of a product, service or business model into the ethical evaluation. Secondly, the increasing use of optimisation technologies highlights the fact that their ethical evaluation is a complex and often inherently political undertaking which can only be answered through societal discourse and public deliberation. Take, for example, the at first sight rather innocuous optimisation of routes of public school buses in Boston (Bertsimas, Delarue, and Martin 2019). It demonstrates that in addition to more efficient bus routes to reduce costs, traffic volume and CO_2 emissions, major health issues and different individual needs of children, e.g. with special needs, must be taken into account. The multiplicity of different variables to be included in an optimisation function poses an immense challenge to achieve a fair result with acceptable trade-offs with which those affected are satisfied (Scharfenberg 2018; Ito 2018; Crockford and Ito 2017).

A focus on the ethical design of AI systems on a technical level thus risks ignoring essential and fundamental aspects. Even if a system is technically mature and meets the highest standards of accuracy, fairness and privacy, its use may be ethically problematic because it overlooks trade-offs or may reinforce structural social injustices as in the case of predictive policing (Green 2020). While ethical aspects at the micro level of the technical system constitute a key element of AI ethics today, crucial business decisions and practices that implement these systems in products, services and business models have been largely neglected. As a result, however, questions relating to, e.g. the concentration of power, practices of attention hacking, or concerning structural injustices such as institutional racism or problematic profit motives are ultimately not being addressed (Greene, Hoffmann, and Stark 2019). Moreover, too narrow a focus not only assumes that ethical challenges arise from flawed or inadequate design of the AI system (Greene, Hoffmann, and Stark 2019; B. Mittelstadt 2019), but also limits the scope of possible options for action in the sense of a technical fix (B. Mittelstadt 2019; Hagendorff 2020; Morley et al. 2020; Greene, Hoffmann, and Stark 2019). The ethical relevance of business practices and the wider societal context shows that a focus on "better building" (Greene, Hoffmann, and Stark 2019) is insufficient as ethical implications go beyond an ethical design of AI as a technical system and AI ethics cannot be "solved" but should rather accompany the use of AI continuously (B. Mittelstadt 2019; Metcalf, Moss, and Boyd 2020). In this light, approaches to "ethics by design" (Dignum et al. 2018; d'Aquin et al. 2018; Aizenberg and van den Hoven 2020) may reveal similar limitations insofar as they are based on the assumption that ethical questions can be dealt with exclusively or predominantly at the level of the design of a system. The implicit assumption of moral causation in the sense that poor ethics on the part of the responsible developers are the source of bad designs which in turn produce harmful outcomes (Greene, Hoffmann, and Stark 2019) reflects at least a limited understanding, in the worst case it indicates more fundamental normative shortcomings. Although the relevance of conflicts between short-term profit interests and truly ethical AI have been recognised (Morley et al. 2020; Hagendorff 2020; Veale 2020; B. Mittelstadt 2019), such aspects often remain outside the current focus on ethical design. However, this should not be seen as imposing an apparently incompatible opposition between business on the one hand and ethics or society on the other (cf. Hagendorff 2020). Nor should it mean that the commercial exploitation of AI is in itself ethically problematic. The point is that due to its narrow focus AI ethics does not include an integral part of AI systems as developed and employed in society without considering business models, business practices, their potential wider impacts and the general societal context which they are

part of. The ultimate danger here is for AI ethics to become ineffective and powerless (Rességuier and Rodrigues 2020; Luciano Floridi 2019; Benkler 2019). In response to this shortcoming, the challenge is therefore to expand AI ethics in such a way that the use and integration of AI systems in business practices and the necessary negotiation of legitimate (optimisation) goals and trade-offs can also adequately be taken into account.

2. AI ethics is biased toward a technological solutionism

Another reported deficiency of current AI ethics lies in the tendency to ignore the question of whether and when the use of AI systems may be less appropriate than another solution (Green 2020; Greene, Hoffmann, and Stark 2019; Metcalf, Moss, and Boyd 2020; Hagendorff 2020; B. Mittelstadt 2019). At least three different elements can be distinguished with respect to this type of technological solutionism. First, following a technically driven perspective, AI ethics seems to take technical progress and the development and use of AI as given and somewhat unchangeable (Green 2020; Greene, Hoffmann, and Stark 2019). Yet this loses sight of the fact that technical progress always takes place within the scope of economic, political and social conditions. To the extent that technical advances are thus always the result of the societal conditions under which they are achieved, they are normatively shaped and not invariable. The development and use of new technologies like AI is therefore always informed by societal values, no matter how hidden they may be. This implicit adoption of technological determinism raises a second element. According to this, technological solutionism leads AI ethics to neglect the question of whether an AI system is in fact the most suitable and effective solution for the problem at hand (Green 2020). The question as to when a (semi-)automated decision-making system is actually the best choice, whether human decisions may be useful in a specific case (Hagendorff 2020) or whether the cause of the problem is not rather to be found on a structural and systemic level (Green 2019; Veale 2020), is of utmost ethical importance. Or as Greene et al. put it (2019, 2127), p. 2127: the "ethical debate is largely limited to appropriate design and implementation—not whether these systems should be built in the first place." Finally, third, this kind of technological solutionism implies restricting AI ethics to technical solutions to address ethical challenges. However, this not only limits the range of possible courses of action and levels at which changes are necessary for ethical AI. It also narrows the view of where and which ethical questions arise at all: when holding a hammer, everything looks like a nail. This tendency of a technical fix in AI ethics thus risks overlooking important ethical questions, curtailing complex,

ethical questions and thus avoiding a wider societal debate (B. Mittelstadt 2019; Greene, Hoffmann, and Stark 2019). But fundamentally, as the examples above reveal, "AI ethics is effectively a microcosm of the political and ethical challenges faced in society" (B. Mittelstadt 2019, 505), p. 505. Recognising this means, among other things, that more emphasis must be placed on the question of the ethical appropriateness of (the use of) an AI system, e.g. in relation to the causes of the problem to be solved and possible (non-)technical alternatives.

3. AI ethics succumbs to an individualist focus

The first two points of criticism are closely linked to the aspect of an individualistic focus. As pointed out in the example of optimisation technologies above, AI ethics mainly examines the ethical implications in relation to individuals, i.e. whether the privacy of persons is sufficiently protected, whether persons are unfairly discriminated against or whether the results of AI systems are sufficiently comprehensible for its users. However, this overlooks ethically relevant effects that the use of AI systems may have on groups or society as a whole (Veale 2020; Morley et al. 2020). While Morley et al. (2020) highlight the role of trust, questions of societal monitoring, control and governance and their political impact, particularly on democratic societies, are often discussed in public debate (Zeng et al. 2010; Yang et al. 2019; Yeung 2017). An overly individualistic focus therefore risks not addressing important ethical consequences at the societal level. In addition, Hagendorff (2020) points to a noticeable omission of more often than not hidden social and ecological costs, such as the outsourcing of necessary labelling of data sets to so-called "clickworkers" or the extensive energy consumption caused by necessary hardware services. While this may be understood as a weakness in relation to the first wave of AI ethics, the problem also persists when it comes to the question of implementation. To the extent that there is a tendency to implement ethics in the sense of "better building" by means of technical solutions, mainly developers and data scientists are assumed responsible for ethical action. In addition to the application of appropriate technical measures, this is reflected, for example, in the development of professional ethics (Bertelsmann Stiftung 2018; Boddington 2017), the teaching of ethics to AI practitioners (Goldsmith and Burton 2017) or tools such as checklists (Madaio et al. 2020; Gebru et al. 2020; Ryan and Stahl 2020) directed at developers and data scientists. But also critical contributions, which rather belong to an emerging third wave of AI ethics, sometimes tend to argue with a focus on individuals as relevant actors for ethical AI (Green 2020). The point is not to say that this perspective on appropriate action would be unjustified or ineffective

– because it certainly is not. Instead, we wish to highlight that this form of an individual focus tends to lose sight of the role of the organisational level, i.e. of businesses, their strategies and business models, but also of questions of internal governance and corporate culture (cf. Metcalf, Moss, and Boyd 2020) as important levers for ethical AI. In the words of Mittelstadt (2019, 505), p. 505: "This approach conveniently steers debate towards the transgressions of unethical individuals, and away from the collective failure of unethical organisations and business models." Consequently, the lack of an individualistic orientation shows two things. On the one hand, a wider deliberative approach is needed to discuss and assess the complex social impacts appropriately. And secondly, the role of organisations and companies as actors should be given more weight as responsibility for ethical action should not be assigned at the individual level alone.

4. AI ethics is problematic in its implementation and lacks accountability and clear impact

The fourth weakness can be summarised as the problem of implementing ethical principles. One reason for this lies in the often very abstract and vague formulation of ethical principles, which leave room for different interpretations (Jobin, Ienca, and Vayena 2019). This results not only in the risk of divergent interpretations, unclear claims and negative effects on trust (Morley et al. 2020), but also in a rather vague basis for attempts of operationalisation and implementation in legal, organisational or technical contexts. Beyond that, the challenge of translating abstract ethical principles into specific requirements may be one reason for the focus on technical solutions, given that technical parameters provide precise specifications for the implementation of ethics. On a social, political, legal, governance or corporate culture level, the field of possible measures and methods of implementation appears to be much more diverse – and hence much more complicated. The abstract formulation of ethical principles thus leads to the considerable difficulty of developing and implementing approaches for their practical implementation (Morley et al. 2020; Jobin, Ienca, and Vayena 2019; Whittlestone et al. 2019; Coeckelbergh 2019; Hagendorff 2020; Luciano Floridi 2019), not least on a legal level (Larsson 2020). Besides suitable tools and measures, this also applies to the definition of responsible actors and accountability structures that ensure that the principles are complied with at all (B. Mittelstadt 2019; Jobin, Ienca, and Vayena 2019). While appropriate approaches to the implementation of AI ethics are urgently needed, their mere existence is not sufficient. Effective structures and robust processes need to be established, evaluated

and documented in order to enable a sustainable impact of AI ethics (B. Mittelstadt 2019). What this shortcoming of AI ethics shows is not only the difficulty of putting ethics into practice, especially in a business context. It also points to the fundamental discrepancy between normative goals and practical approaches, often due to the lack of an explicit and theoretically sound normative framework (Green 2020) to justify both particular normative goals and the means of their effective implementation. As a consequence, a third wave of AI ethics should focus on substantiating normative goals based on a solid theoretical foundation in order to derive practical approaches and counteract a gap between formulated principles and their practical implementation.

5. AI ethics lacks a clear relationship to legal regulation

Finally, a fifth weakness can be summarised as the often unclear relationship to the legal regulation of AI systems. The dynamism with which the first and second wave of AI ethics were triggered and large technology corporations dominated the resulting public discourse led to concerns that industry could determine the ethical standards to be applied to AI (Benkler 2019). Although the concern is closely linked to the economic power of many large corporations, the often conceptually ambiguous relationship of AI ethics to pivotal legal issues, such as the impact of AI on existing legislation or the need for further legal regulation, contributes to this concern. Rességuier and Rodrigues (2020) argue that this is due to an underlying law conception of ethics which misunderstands the role of ethical principles and thus risks the practical effectiveness of AI ethics. Beyond mere virtual signalling, the ambiguous use of "ethics" on a communicative level may be tactically exploited to influence the public debate and prospective legislation. An undefined relation of ethics and law thus risks AI ethics being misused to soften, delay or prevent hard legal regulations (B. Mittelstadt 2019; Jobin, Ienca, and Vayena 2019; Green 2020; Wagner 2018; Coeckelbergh 2019; Veale 2020). Moreover, even at the political level, the relationship between ethics and law sometimes seems to be unclear in the context of AI (Wagner 2018), which may become problematic in view of the time delay in legislative processes reacting to rapid technological developments such as AI (Larsson 2020). Although the danger of "ethics washing" thus seems reasonable (Benkler 2019; Green 2020; Rességuier and Rodrigues 2020; Tsamados et al. 2020; Veale 2020; Larsson 2020; Metcalf, Moss, and Boyd 2020; Morley et al. 2020), it is important to note that from a conceptual perspective the relationship of ethics to questions of legal regulation can be considered complementary in principle (Luciano Floridi 2018). In contrast to legal legislation, ethics is particularly helpful when legislation is unavailable, requires ethical

interpretation or counterbalance, or when something that is (still) legal should be avoided for ethical reasons, or something not yet legally required should be done for ethical reasons (Luciano Floridi 2019). Even though AI ethics lobbying, that is “the malpractice of exploiting digital ethics to delay, revise, replace, or avoid good and necessary legislation (or its enforcement) about the design, development, and deployment of digital processes, products, services, or other solutions” (Luciano Floridi 2019, 188), p. 188, poses a significant risk of undermining serious ethical efforts, a conceptually clear demarcation from questions concerning the legal containment of AI systems is possible. In short, we conclude that the fifth identified shortcoming requires a third wave of AI ethics to clearly determine its legitimate role and promote appropriate communication activities. Building on a solid normative foundation, AI ethics should thus describe both its tasks and limitations.

As this review of the evolution of AI ethics demonstrates, at the beginning of a third wave some key steps need to be taken to ensure that AI ethics can make an effective long-term contribution to technology, the economy and society. Based on the principles-led approaches of the first and manifold efforts for the practical implementation of the second wave, ethical implications of AI-based business models and business practices on a societal level need to be brought more into focus. In addition to "better building" (Greene, Hoffmann, and Stark 2019) the goal of “better managing”, in the sense of considering the wider social, economic and ecological consequences, needs to become a key element of AI ethics. The call for a transition to "microethics" (Hagendorff 2020; cf. L. Floridi 1999) should therefore be complemented by a perspective of "macroethics", which deals with the ethics of products and services at the level of markets and the organisational relationship between businesses and society as a whole. Secondly, this includes to extend current approaches that take the development and deployment of AI systems as a given and irreversible fact and concentrate on technological answers, so that the deployment of AI systems as such can be reflected and wider options for action are enabled. Thirdly, a third wave of AI ethics needs not only to take greater account of the wider impacts on societies but also focus on businesses at an organisational level as responsible actors for ethical behaviour. Future approaches to AI ethics should finally adopt conceptually clear and transparent demarcation of the legal regulation of AI and openly address challenges in implementing ethics. Besides practical approaches for everyday business, the implementation of AI ethics should also comprise issues of effectiveness, accountability, and the justification of both normative goals and proposed measures. In conclusion, the five concerns as described above point to the weakness of current

AI ethics in recognising fundamental normative challenges and acknowledging the inherent political dimension of AI ethics (B. Mittelstadt 2019; Greene, Hoffmann, and Stark 2019; Green 2020; Wong 2020). This manifests itself in the neglect of the business context, a strong focus on ethical design and a primary attribution of responsibility to individuals such as developers. Against this background, we aim in the following chapter to offer a first step towards complementing AI ethics by drawing on established normative theories from business ethics.

4.2.5 Order ethics as business ethics approach to AI

In this chapter, we present a contractualist theory of business ethics arguing that it provides a suitable normative approach to AI ethics. Since a comprehensive introduction to the philosophical foundations of contractualism or business ethics is beyond the scope of this article, we focus on the aspects essential to our reasoning. We first introduce the concept of order ethics and then contrast it with integrative social contracts theory (ISCT) as the most prominent example of contractualist business ethics.

Business ethics deals with the question of the possibility of ethical behaviour in a market economy which is driven by the principle of competition (Moriarty 2017). Despite early contributions on AI from a business ethics perspective (Khalil 1993) and the fact that the impact of AI on business ethics has been recognised (Kaplan and Haenlein 2020, 2019) and conceptualised by several authors (K. Martin, Shilton, and Smith 2019; K. Martin 2019; Ryan and Stahl 2020; Bartneck et al. 2021), business ethics approaches are hardly found in the current AI ethics debate (Lütge 2019; K. Martin, Shilton, and Smith 2019), but in no case from a contractualist perspective.

As a concept of contractualist business ethics, order ethics refers to constitutional versions of contractualist theory (J. M. Buchanan 1975; G. Brennan and Buchanan 1985) which provide for the fundamental attribution of basic rights, e.g. based on human rights, via a constitutional contract, and thus go beyond more reductionist approaches of contractualism building on J. Locke or R. Nozick. Although to some extent similar to J. Rawls' contractualist *Theory of Justice* (1971), two key distinctions can be made (Luetge, Armbrüster, and Müller 2016): first, the negotiation of contractual conditions does not take place in an idealised setting behind a veil of ignorance but is shifted to the real-world situation of business ethics. Secondly, a constitutional version of contractualism does not seek to derive normative principles that determine a just social order but reflects solely on the normative foundations of economic action based on the assumption of self-interested persons. In this sense, it is an economic approach to business ethics that promises to be particularly compatible with business practice.

The starting point of order ethics are value conflicts which are addressed and aimed to be resolved from a contractualist perspective. As will be shown, it is this fundamental approach that makes order ethics a promising complement to AI ethics. Confronted with conflicts between different values – ranging from individual interests, to social norms or ethical values – order ethics assumes that no recourse to a certain substantial

normative principles is possible, however they may be defined and justified in advance in any form, but that a solution can only be reached by agreeing to a rule for the benefit and in the interest of all parties involved. Ethical conflicts in this sense are to be negotiated and resolved only through a solution that settles the conflict in the sense of a voluntary agreement on the basis of individual consent, but not through reference to higher normative principles. In this sense, the contractualism of order ethics is both more and less ambitious than Rawls' understanding of it (Luetge, Armbrüster, and Müller 2016): less ambitious, as no attempt is made to justify overarching normative principles, and more ambitious, as this means that contractual renegotiation in the face of ethical conflicts takes place under real-world conditions with all the associated entanglements and complications. For order ethics the level on which agreements are made is essential. Based on the distinction between action and rules (G. Brennan and Buchanan 1985), order ethics holds that ethical conflicts can only be resolved in a justifiable manner at the level of the conditions for action. In this way, order ethics responds to the risk of ethical behaviour being crowded out, since more often than not it is not rewarded at the level of individual actions in a competitive environment (Luetge 2013). Typically conceptualised in the form of the prisoner's dilemma, order ethics thus reflects the problem of cooperative behaviour (in competitive markets): only if ethical standards are set at the level of rules can individual ethical behaviour be reasonably required since otherwise they will be subject to some form of sanction. With this in view order ethics advocates the following concerning the notion of rules (Luetge, Armbrüster, and Müller 2016, 692; citing G. Brennan and Buchanan 1985), p. 692:

[1] Only changes in rules can change the situation for all participants involved at the same time.

[2] Only rules can be enforced by sanctions – which alone can change the incentives in a lasting way.

[3] Only by incorporating ethical ideas in (incentive-compatible) rules can competition be made productive, making individuals' moves morally autonomous in principle. With the aid of rules, of adequate conditions of actions, competition can realise advantages for all people involved.

First of all, rules need not be understood in a narrow economic or political sense as they can also be drawn from ideas from other areas of society such as culture, philosophy or arts (Luetge, Armbrüster, and Müller 2016). What is further important is that corresponding rule changes or new rules designed to resolve ethical conflicts do not conflict with individual actions, so that no counteracting incentives on the level of

rules arise (Luetge 2016). The shift from ethics to the level of rules means that ethical conflicts should be clarified by deriving more general rules that apply not only to the specific individual case at hand but to at least one specific group of conflicts and actors. In this sense, it is about finding rules of distribution of goods and not about determining one particular distribution of goods (Luetge, Armbrüster, and Müller 2016). Not least, an agreement on the level of rules facilitates the consent of all parties involved. Although order ethics thus underlines the importance of an appropriate general framework for ethical behaviour through the concept of rules, rules should not be put into one with laws. Instead, order ethics seeks to provide a conceptual supplement to laws and the general legal framework based on the theory of incomplete contracts (Hart 2017) which may also raise ethical conflicts. Incomplete contracts occur, for example, when obligations are not sufficiently clarified, when it is difficult to assess whether a contract has been complied with or when its enforcement is difficult (Luetge 2016). To the extent that it is impossible to adequately equip all contracts for all possible future scenarios and to amend incomplete contracts, their occurrence is necessary and cannot be avoided. The resulting scope for interpretation of legal contracts, which deal with complex issues or claim validity over a long period of time, should thus not be seen as a shortcoming but rather as an advantage in dynamic environments by allowing flexibility and adaptability. Order ethics understands the role of ethics in managing the openness of incomplete contracts, including the resulting uncertainty and possibly emerging conflicts (Luetge 2005; Luetge, Armbrüster, and Müller 2016; Luetge 2016). This allows order ethics to define the place of ethics and to specify its relationship to legal regulations. "Order" therefore does not refer to the legal framework but to all other formal and informal rules and agreements which seek to enforce ethical behaviour, for example at the level of individual sectors or groups of firms (Lütge 2019).

By shifting the focus of ethics to the level of rules, order ethics finally emphasises the contractualist criterion of mutual benefits (Heugens, van Oosterhout, and Kaptein 2006; Luetge 2005; Luetge, Armbrüster, and Müller 2016). Accordingly, given the absence of overarching normative principles, only such an agreement can be normatively justified which offers benefits for each individual or party involved on the basis of his or her individual values and interests. In this context, possible advantages are to be understood broadly and include not only monetary or financial benefits, but everything that people take to be advantages (Luetge 2005). In practical terms, firms should resolve ethical conflicts arising, for example, from previously missing, impossible or unintentional legal regulations by means of adapted or new rules, which

are in the interest of every stakeholder involved and thus generate mutual benefits. For only when real win-win situations are created (Luetge, Armbrüster, and Müller 2016) a normatively justified solution can be claimed. This does not imply that firms should abandon a business management perspective but rather that they must improve their economic calculations by incorporating the values of various stakeholders and, for example, taking into account long-term effects on reputation (Luetge 2005).

In a nutshell the core elements of order ethics can be summarised as follows:

[1] Building on contractualism as normative theory, order ethics argues that ethical conflicts cannot be resolved by reference to overarching normative principles (reasonable pluralism).
[2] Instead, ethical conflicts ought to be solved by adapted or new rules to which each stakeholder involved consents based on their individual values.
[3] The normative criterion is the mutual advantage that is to be achieved by a respective agreement.

We conclude the brief introduction of the concept of order ethics by highlighting some of its main advantages. First, the concept of order ethics is rooted in a fundamentally pluralistic view of society. According to this, a multitude of different values can be legitimately held, which ultimately may come into conflict with each other. In resolving these conflicts, no shared basis of common values of any kind should be assumed but rather each individual value is accepted as normatively justifiable. This offers a key advantage over other ethical theories of business ethics such as utilitarian approaches. Roughly speaking, the latter assume that in the face of an ethical conflict, the option that yields the greatest possible (measurable) benefit should be chosen. However, this not only bears the risk of delivering highly counterintuitive results but more importantly it requires the maximisation of utility, however defined and justified as universal ethical norm. Secondly, a contractualist approach seems to be better equipped than stakeholder theories of business ethics to reconcile claims of different stakeholders, balance incommensurable conflicts of values or solve problems of collective action, given that contract theories were originally formulated to address these very issues (Wempe 2004). Third, by aiming at rule changes, crucial constraints of operating in a competitive environment can be taken into account. For as Morley et al. (2020, 2161), p. 2161, note, it is highly plausible that not least in the context of AI, an ethical approach would constitute a competitive disadvantage for any single "first mover". Fourth, order ethics allows us to specify the relationship between ethics and legislation, which is of particular relevance to AI ethics. Before we discuss the

implications in more detail at the end of this chapter, we first contrast the introduced concept of order ethics with the probably most prolific theory of contractualist business ethics, namely integrative social contracts theory (ISCT).

4.2.5.1 A cursory comparison of two contractualist theories of business ethics: order ethics and ISCT

Just as order ethics, integrative social contracts theory (ISCT), originally developed by T. Donaldson and T. W. Dunfee (1999; 1994, 1995), stems from a contractualist basis. For ISCT too, the central question is how conflicts between different or differently prioritised values and norms can be overcome. In contrast to order ethics, however, ISCT assumes a macrosocial contract which sets the conditions for microsocial contracts. Although Donaldson and Dunfee do not assume a strong hypothetical setting in the sense of Rawls' veil of ignorance as the contractors know at least their basic preferences and values, they nevertheless assume that "information about their personal economic endowments and roles in society" (Dunfee and Donaldson 2015) is unknown. Under the four terms of the macrosocial contract, the following conditions are set out (Dunfee and Donaldson 2015; T. Donaldson and Dunfee 1994; T. J. Donaldson and Dunfee 1999):

[1] Local communities may specify ethical norms for their members through micro-social contracts (called "moral free space").
[2] Norm-generating microsocial contracts must be grounded in informed consent buttressed by a right of community members to exit and to exercise voice within their communities.
[3] In order to be obligatory (legitimate), a microsocial contract must be compatible with hypernorms.
[4] In case of conflicts among norms satisfying principles 1–3, priority must be established through the application of rules consistent with the spirit and letter of the macrosocial contract.

According to ISCT, the actual discussion of ethical conflicts is moved to the level of micro-social contracts, which all members of a local community must agree to for the agreement to be considered an authentic norm. The members of a community have the right to leave the agreement and to give voice to their position. Individuals may be members of several economic communities, defined as "self-circumscribed group of people who interact in the context of shared tasks, values, or goals and who are capable of establishing norms of ethical behaviour for themselves" (T. Donaldson and Dunfee

1994, 262), p. 262. Decisive for the legitimacy of micro-social contracts is their compliance with so-called hypernorms, certain universal ethical principles such as those expressed in human rights (T. Donaldson and Dunfee 1994; T. J. Donaldson and Dunfee 1999). These can be either procedural hypernorms such as the right to exit and voice, substantive hypernorms such as respect for human dignity or structural hypernorms such as the right to property or necessary social efficiency (Dunfee and Donaldson 2015). Lastly, ISCT stipulates that conflicts between microsocial norms will be resolved by so-called priority rules, provided that they are in line with hypernorms. A total of six such rules decide how to deal with conflicts in case of doubt. A large part of the practical implementation of ISCT, besides the identification of stakeholders of a community, rests thus in the empirical determination of microsocial authentic norms (Dunfee 1991) and the identification of relevant hypernorms. Ever since its original introduction in the mid-1990s, ISCT was criticised (e.g. Wempe 2004, 2008; Soule 2002; Phillips and Johnson-Cramer 2006; Boatright 2000; Calton 2006; Reisel and Sama 2003; Van Buren III 2001) and defended (Dunfee and Donaldson 2015; Dunfee 2006) and has become an important cornerstone in the debate on contractualist business ethics.

Starting from this rough summary of ISCT, some similarities and differences to the theory of order ethics can be noted, given they both provide approaches of contractualist business ethics. Fundamentally, both concepts bear similarities in their contractualist foundation, according to which they start from two different levels, constitutional and post-constitutional rules based on Buchanan (1975) in the case of order ethics and macro and microsocial contracts in the case of ISCT. Furthermore, against the backdrop of their similar theoretical framework, both approaches emphasise the role of individual consent for the legitimate validity of rules or authentic norms, with ISCT particularly emphasising the role of consent in the sense of engaging in a practice (T. Donaldson and Dunfee 1994). Beyond these underlying similarities, however, significant differences in the way the concepts are further elaborated can be identified. Most noticeable appears to be the handling of ethical conflicts. While order ethics seeks to resolve conflicts through mutually advantageous rule changes, ISCT establishes hypernorms, i.e. universally justified principles (Luetge, Armbrüster, and Müller 2016). Referring to third normative principles, however, creates serious problems of justifiability, legitimacy and empirical identification given the assumption of reasonable pluralism. Also, in practice, the identification of hypernorms seems to result in a much more complicated process as it involves high justification standards. In this light, the concept of order ethics seems to be more suitable for the context of

AI as it works on the grounds of weaker normative requirements. The second difference we notice concerns the perspective from which ethical conflicts are approached. ISCT considers these conflicts to be deficiencies of a market economy and that they should be corrected accordingly (Luetge, Armbrüster, and Müller 2016). Not least, this limits the scope of possible outcomes of ISCT to standards like code of conducts and results in a mechanical approach to business ethics (Burg 2009), which, as we have shown above, is not adequate for the context of AI. Order ethics, on the other hand, aims to achieve mutual benefits for all stakeholders of an ethical conflict through rule changes and in this sense, it strives to ethically improve the market economic system. Again, order ethics seems to offer a better approach for the context of AI. Because especially in a dynamic and rapidly developing field of technology, it is important to actively shape innovations through ethics. This type of productive perspective is facilitated by an opportunity-oriented approach rather than an approach geared to remedying deficiencies. Moreover, solving emerging ethical conflicts through a set of six priority rules seems to present a somewhat rigid (Calton 2006; Phillips and Johnson-Cramer 2006) and probably conservative (Husted 1999) framework, which appears to be ill-suited for the dynamic context of AI.

Nevertheless, we would like to point out one aspect which we think is worth being added to order ethics from the concept of ISCT in the context of AI. This concerns the characterisation of economic communities as respective subjects of ethical decisions. We find that this conceptualisation fits particularly well into the concept of order ethics as it provides a suitable starting point for its procedural expansion in the context of AI. In the next section we will argue for a procedural amendment of order ethics providing a practical method to deal with ethical conflicts between values and interests in the context of AI.

4.2.6 Community-in-the loop: The concept of deliberative order ethics

4.2.6.1 Bringing business ethics to AI: A procedural extension of order ethics

In the following section we will introduce order ethics as a theory of normative business ethics to the field of AI. To this end, we advocate that order ethics provides a suitable framework of normative business ethics to complement AI ethics as presented in the first part of the paper. However, we also argue for a procedural addition through deliberative stakeholder engagement that provides a suitable methodological extension to debate value conflicts and agree on trade-offs via adequate rules.

The starting point of order ethics is the question of how to deal with ethical conflicts that may arise for firms given the competitive environment of international market economies. We believe that this approach to ethics provides a valuable addition to the predominant perspective of current AI ethics considering the shortcomings as identified above. As Wempe (2009) explains, ethical conflicts between different norms and values may arise due to globalisation, increasing complexity, increasing specialisation. This applies especially to the context of AI. Importantly, the perspective of ethical conflicts allows issues beyond ethical design to be brought into the focus. Because besides conflicts between accuracy, accountability or fairness (Corbett-Davies et al. 2017), in particular conflicts between very diverse and complex issues have to be taken into account when assessing an AI system. Ultimately, the assessment of conflicts between different values, norms or interests is about determining the necessary trade-offs and negotiating which solution and distribution of costs and benefits is acceptable for all parties involved. Some of these conflicts are already inherent in the concepts currently employed by AI ethics, such as fairness or privacy, the application of which therefore requires a thorough normative analysis (Kleinberg, Mullainathan, and Raghavan 2017; Kleinberg et al. 2018; T. Miller 2019; Wong 2020; Binns 2020; B. Mittelstadt 2019). Other potential trade-offs include, for example, those between the intended purpose of an AI system and resulting costs for employees in terms of layoffs or training, costs for suppliers or other partners in terms of systemic risks or resulting dependencies, complex social or ethical costs in terms of gains in flexibility, risks to surveillance and privacy, direct or indirect costs to society through monopolisation effects or beneficial alternatives that are being pushed aside, or costs for the environment from energy consumption or the mining of raw materials. While some of these conflicts and trade-offs may be explicitly considered and perhaps even included in the cost calculation of a system and business model, such as the risks of

safety and security, others, especially unintended and longer-term consequences, are often difficult to identify at all (Rahwan 2018). Take the example of the above-mentioned recruiting systems. What effects does the increased use of AI-based recruiting systems have on applicants, on the human resource management in firms and on the labour market in general? Under what conditions does their use seem acceptable to all stakeholders in the long term? The example of optimisation technologies illustrates the complexity of the conflicts: how should benefits and drawbacks for children, parents, teachers, schools, public administration and bus companies in terms of health effects, cost and time savings and environmental effects be best organised for all stakeholders? Along similar lines, Whittlestone et al. (2019) describe such conflicts as tensions with which AI ethics is confronted. By summarising four such key tensions in general terms, they highlight the challenge of assessing costs and benefits. It becomes clear that the identification and judgement of such value conflicts is a political task by its very nature, which involves the social negotiation of different values, conflicts and trade-offs (Whittlestone et al. 2019).

Insofar as the contractualist theory of order ethics starts out from precisely such ethical conflicts, the approach seems particularly apt to complement AI ethics at this point. As order ethics is based on reasonable pluralism respecting the multitude of values that prevail in society, no substantial basis in the sense of a certain set of shared values is assumed. For the context of AI this means that all values and interests must be given equal consideration in emerging conflicts, without any of them being in any way given a lower valuation than others. Nor would it be possible to reduce conflicting values to some kind of common basic value. No matter how great a challenge this presents for order ethics, it is essential to recognise the pluralism of values. Order ethics now provides for agreements on the level of rules to which the stakeholders involved agree on the basis of the normative criterion of mutual advantage (Luetge 2005). At this point, we propose to add an important procedural element to order ethics in order to develop and agree by means of participation and deliberation on a suitable measure at rule level, in which all stakeholders can realise their values, i.e. achieve benefits of some kind. Our proposal thus amounts to the following: in order to deal with conflicts between different values and to arrive at an assessment of trade-offs and a fair distribution of costs and benefits of an AI-based product or service, stakeholders responsible for or affected by a business practice should formulate a rule through a deliberative participation process to which all can agree on the basis of their own interests. We argue that participation is the appropriate method for deriving an eligible rule (or set of rules) since the legitimate interest of stakeholders is already manifested

in the criterion of mutual advantages and the collaborative, co-creative development of an eligible rule is therefore the most effective way to meet it. To negotiate complex value conflicts in the context of AI, the participatory involvement of stakeholders as well as the cooperative consultation is necessary since only in this way relevant values and interests as well as diverse consequences and benefits and costs can be identified, and ultimately legitimate trade-offs balanced. First, only in this way can the diverse values and potential costs be determined because for most of them there are neither any validated data or parameters nor standards for their evaluation. Second, only a deliberative process allows decisions to be made on whether trade-offs are acceptable and whether the balance between advantages and disadvantages is societally desirable. The question of which social groups (e.g. children, children with special needs, teachers, bus drivers, etc.) should benefit or bear which disadvantages, how health effects should be weighed against cost savings and environmental improvements are complex societal negotiation processes. Similarly, in the case of AI systems in recruiting, questions may arise such as how to reconcile efficiency gains for firms with potential benefits and harms for certain groups of applicants, potentially increasing dissatisfaction and emotional distress for applicants, or with increased insecurity in labour markets. Deliberating about the different costs and benefits for respective stakeholders is therefore a suitable approach to prioritise values, decide on trade-offs and thus do justice to the political dimension of the problem. Only through participatory and deliberative exchange can a societal consensus and, building on this, an agreement be found which provides acceptable benefits for all. In our view, participation and deliberation are the appropriate methodological strategies to make AI ethics, within the framework of order ethics, a societal and political debate on the consequences of AI-based business practices at the level of organisations and actors. With the procedural supplement to order ethics presented here, we hope to adapt the crucial step of rule changes to the context of AI ethics.

One important element is the question of legitimate stakeholders. For order ethics it is central to develop ideas on the level of rules, so that ethical behaviour does not cause a competitive disadvantage for individual actors. Depending on the individually defined scope, stakeholders may be, for example, those who belong to a specific industry or a specific area of application of AI systems, such as AI in recruiting or human resource management, or AI in the public sector or for public infrastructures. Stakeholders include those involved in the development and employment of AI systems as well as those potentially affected, in particular specific groups from civil society. Ultimately, the identification of relevant stakeholders depends on the precise

definition of the specific scope that the rule to be developed should cover. It is likely, however, that this can only be finally determined in the participatory deliberation process itself, as it is often anything but trivial to decide at which level a rule is effective and compatible with competition. For this purpose, we suggest borrowing the term community from ISCT (T. J. Donaldson and Dunfee 1999; T. Donaldson and Dunfee 1994) to describe as an economic community a group of stakeholders who are interested in the ethical governance of AI systems on the basis of a shared interest in a specific field of application.

Furthermore, the issue of rules is essential. In general, rules can be drawn from a wide variety of conceptual ideas and therefore do not need to be legitimised by a specific legal, political or economic background (Luetge, Armbrüster, and Müller 2016). Rather, the aim is to give voice to the pluralism and capabilities of deliberative participation processes through creative rules. The only requirement is that beyond resolving one individual case and assessing the costs, benefits and trade-offs of a concrete AI system, the rules must apply to at least a certain group of corresponding products, services or AI-based business practices. With regard to the examples consulted, this might include rules for AI-based business practices in recruiting or human resource management.

The participatory and deliberative extension of order ethics can be further explored in the light of some critiques of ISCT, which argue that its rather static approach is not sufficiently equipped for dynamic contexts of changing norms and conflicts (Phillips and Johnson-Cramer 2006; Calton 2006; Burg 2009; Ast 2019). Burg (2009), for instance, analyses ISCT's concept of authentic norms and criticises Donaldson and Dunfee's recurrent recommendation of corporate codes as an appropriate measure. According to him, this form of "mechanical business ethics" seems problematic: “At their best, codes are merely levers for internal and external stakeholders to hold organisations and organisational actors accountable by stating what is obvious to nearly everyone. At their worst, codes present an ethical façade that is only marginally related to manifest organisational norms, to be treated as the punch line of a joke about how one should behave within an organisation (‘Check the code of conduct!’)” (Burg 2009, 675), p. 675. Not least, this point is reminiscent of the problems of a too principled approach to AI ethics described above. Alternatively, Burg advocates an approach of deliberative business ethics which establishes and prioritises norms by an open process of stakeholder dialogue and ultimately reaches agreements based on consent. Similarly, Phillips and Johnson-Cramer (2006) have criticised the lack of dynamism in ISCT arguing that the described mechanisms of exit and voice do not adequately

reflect the dynamic processes of norm evolution. For a dynamic addition to ISCT they propose four principles, including the principle of community discourse "to create systems for the exercise of voice" (Phillips and Johnson-Cramer 2006, 298), p. 298. Calton (2006) also formulates a more dynamic and process-oriented supplement to Donaldson and Dunfee's ISCT. According to him, ISCT's reference to hypernorms and the defined priority rules are too inflexible and thus unsuitable to deal with the manifold and dynamic value conflicts in a pluralistic context. He introduces a dialogic twist, allowing stakeholders to find a fair agreement in an interactive learning process. Such a dynamic dialogue process is able "to unleash the full reflective potential of a social contracting theory of business ethics" (Calton 2006, 344), p. 344. Overall, it can be noted that ISCT, as the most advanced theory of contractual business ethics, has already been enriched by various participatory and deliberative approaches (Van Buren III 2001; Reisel and Sama 2003; Dunfee 2006; Wempe 2004). We argue that the advantages of such an extension can also be applied to order ethics in the context of AI. However, since the focus of this article is not on a conceptual extension of order ethics, some essential issues have to remain outstanding. What remains to be clarified, for example, are the specific criteria for the identification of stakeholders (cf. Phillips and Johnson-Cramer 2006), whether particular types of rules may be differentiated and what requirements for consent may be derived. Here it can at least be stated that consent in the sense of an ongoing collaborative process (Burg 2009; cf. Wong 2020) would not only enable a constant monitoring and adjustment of rules for rapidly changing business practices, but could also play an important role in terms of accountability. Furthermore, the challenges of process design and the different starting conditions must also eventually be addressed. In particular, standards must be set that adequately take into account the heterogeneity of the stakeholders involved. How can different levels of knowledge, power imbalances and different cultures and languages be managed in such a way that a fair deliberation process is possible? While it is clearly worth building on prior work from related fields, future research would need to further specify the participatory deliberation process and relevant criteria.

Since a more comprehensive explanation of the proposal goes beyond the scope of this article, we summarise our reasoning as follows:

[1] When introducing AI-based business models, **conflicts** between different values, norms and interests may arise over the distribution of benefits and costs of deploying AI.

[2] In order to decide on a societally desirable distribution of costs and benefits and agree on acceptable trade-offs, the **deliberative participation** of the relevant economic community is necessary.

[3] Through engaging in a participatory process of deliberation, the economic community, i.e. stakeholders of using AI in a particular field of application, ultimately establishes overarching **rules** that enable ethical behaviour without creating competitive disadvantages.

[4] Assuming the same legitimacy of the different interests and values, the decisive normative criterion is that all stakeholders of the community agree to the rules on the basis of **mutual advantages**.

[5] As a result, rules are to be created through inclusive deliberation of the economic communities, enabling ethical AI business practices in the sense of **pluralistic value creation**.

In other words, our proposal is to complement the third wave of AI ethics with a stakeholder engagement approach, according to which, whenever conflicts arise between different values, firms engage in a participatory and deliberative process with the relevant economic community to develop rules that enable ethical behaviour in the field of given business practices. Stakeholders of an economic community refers to all such parties who are involved in any way related to the use of AI systems in a specific field of application, whether as developers, users, or affected person or group of civil society. What is crucial here is the theory-based normative criterion of mutual benefits on the basis of which stakeholders consent to an agreement. Only if all stakeholders recognise satisfactory gains in the ratio of costs and benefits, and in this sense win-win situations are created, may a rule legitimately claim validity and be considered as enabling ethical business practices. While in the case of public school bus services, it seems rather intuitive that all parties involved should benefit from the introduction of an AI-based system, this becomes even more complex in the case of AI-based recruiting systems. What follows is that not only, say, developers and firms as users but also potential applicants must benefit from the use of the respective systems. What is thus characteristic is the aim of creating shared value instead of unilateral business value in terms of financial profits for firms involved. This is in line with Schormair and Gilbert (2020) who present a framework for creating shared value in situations of value conflicts among stakeholders. Comparing approaches of agonistic and deliberative stakeholder engagement, they argue for an integrative approach based on a process of discursive justification. Recognising stakeholder value pluralism, they develop a five-step procedural framework which helps to resolve value conflicts by

steps of discursive sharing and potentially leads to pluralistic stakeholder value creation. It is worth pointing out that both approaches do not attempt to resolve the problem of value conflicts by referring to monistic normative theories or consensus in the sense of an agreement on single values, but rather seek to realise different values and thus mutual benefits within a procedural framework. For AI ethics this means that it should not (only) be about “solving” ethical problems but also the creation of more comprehensive benefits or even the promotion of the common good. The perspective of pluralistic value creation helps to establish "AI for good" (Luciano Floridi et al. 2020) not as a sub-field but as the core prospect of AI ethics.

4.2.7 Discussion: Contractualism, deliberation and AI ethics

In the recent debate on AI ethics, a few different contractualist or deliberative ideas have been put forward (Rahwan 2018; Binns 2018; Wong 2020; Whittlestone et al. 2019; Rosenbaum and Fichman 2019)[11]. In the following section, we discuss some of these proposals in order to refine the approach of deliberative order ethics as outlined above.

Most prominently, Rahwan (2018) has introduced social contract theory to the AI ethics debate by arguing for a conceptual framework of society-in-the-loop (SITL). Based on the paradigm of human-in-the-loop (HITL), he applies the idea that at some point of the algorithmic system a human is involved to provide monitoring and supervisory functions, and adapts it to a more general societal level. As Rahwan (2018, 7), p. 7, puts it: "While HITL AI is about embedding the judgement of individual humans or groups in the optimisation of AI systems with narrow impact, SITL is about embedding the values of society, as a whole, in the algorithmic governance of societal outcomes that have broad implications." Recognising the ethical and societal implications AI systems may have, he makes use of social contract theories in a broader tradition referring to Hobbes, Locke and Rousseau as well as to Rawls and Gauthier in modern times as an adequate framework to deal with fundamental value conflicts and the question to find fair distributions of costs and benefits and acceptable trade-offs. As societies today become increasingly governed by AI-based algorithms, SITL seeks to expand the general social contract to the realm of algorithmic and AI-assisted decision making. In order to be able to agree on acceptable trade-offs, both quantifying externalities as well as ways to articulate values and societal expectations are needed to evaluate AI systems. Rahwan then discusses different methods and techniques that could be used to "bridge the society-in-the-loop gap", from value-sensitive design to crowdsourcing and data-driven tools such as computational social choice, as well as deliberation between stakeholders and public engagement (Rahwan 2018, 10–11), pp. 10-11. While he is not proposing one specific methodological avenue to resolve value conflicts and agree on trade-offs, he nevertheless seems to recognise the significance of public engagement as not only experts alone can decide on societal values, but it is precisely through interaction and deliberation that values and norms emerge and can eventually be agreed upon.

[11] Contractualist arguments in the context of AI have also been put forward on a technical level or looking at specific ethical challenges, such as justifiability (Loi, Ferrario, and Viganò 2020) or autonomous vehicles (Leben 2017). But also from a perspective of discourse ethics criteria of rational communication have been developed to manage algorithmic accountability (Buhmann, Paßmann, and Fieseler 2020).

In general, Rahwan's and our approaches presented above have much in common as they both seek to provide an ethical framework for AI based on the normative theory of contractualism. Against this background, our account of deliberative order ethics is similar to Rahwan's SITL in the conceptualisation of AI ethics as conflicts between different values and interests with regard to the trade-offs AI systems may imply. While both approaches use contractualist theory to address necessary trade-offs, we focus on stakeholder participation and deliberation as appropriate method to find an agreement. Rahwan on the other hand seems to remain rather open in this regard yet emphasises the need for quantifiable tools to measure human values (Rahwan 2018, 9), p. 9. Although we agree that this would help streamline negotiation processes, we remain somewhat sceptical about quantifiable parameters as prerequisite in order to be considered. As the pluralistic approach is explicitly acknowledging the equal authority for any value and interest that stakeholders may hold, thus including economic as well as broad social or ecologic values, we think that the quantifiability condition risks excluding and disadvantaging some values and thus unjustifiably reduces the deliberative arena. Perhaps the most important difference, however, lies in the different levels at which the approaches are ultimately intended to have practical effect. Whereas Rahwan defines the SITL framework in contrast to the HITL paradigm that operates on a micro-level of individual technical systems; our approach focuses on the organisational level of firms based on contractualist business ethics. The SITL thus "looks more like public feedback on regulations and legislations than feedback on frequent micro-level decisions" (Rahwan 2018, 12), p. 12. Deliberative order ethics by contrast seeks to enable ethical behaviour of firms at an intermediate level, below the level of legislation and above individual measures of corporate governance. As pointed out above, the rules which deliberative order ethics aim to establish in order for ethical behaviour not to constitute a competitive disadvantage can be understood as soft law. In Rahwan's framework, negotiation and public engagement would instead take place at the regulative level of hard law. While both our approaches thus acknowledge the crucial political dimension of AI ethics, deliberative order ethics supplements the picture by adding a political and discursive level between legislation and the micro level of technology. It is in this sense that we call our approach *community-in-the-loop* (CITL) adapting Rahwan's reasoning and based on our definition of economic community above. We argue that in general society expresses values and preferences through the political system in place, assuming they are democratic societies. We agree with Rahwan in that more adequate hard legislation in the context of AI is needed and that more participation and deliberation on the level of democratic political systems would be of great value, yet

for ethical AI based on participation to be most effective, we believe an approach of *community-in-the-loop* would be appropriate.

In the context of algorithmic accountability, Binns (2018) proposed a concept based on the democratic ideal of public reason. Binns explores the question of algorithmic accountability, i.e. the right of individuals to know what principles and considerations lie behind an algorithm-based decision in order to be able to understand and, if necessary, contest it (Binns 2018, 547), p. 547. The challenge of algorithmic accountability, according to Binns, is therefore to make the implicit values of technical systems understandable and justifiable in such a way that they might persist in a pluralistic environment. Thus, the task is not only to identify the epistemic and normative assumptions inscribed in the development and design of technical systems, but to provide explanations and justifications that are acceptable to all (potentially) affected individuals. However, in a society in which individuals legitimately hold differing values, a divergence not only in epistemic but particularly in normative standards seems likely. Thus, in order for algorithmic accountability to promote the legitimacy of algorithmic-based decisions, Binns suggests that an account of public reason be taken as a basis. Put simply, this states that despite existing differences, there must be universal rules and principles "provided they are suitably public and shared by all reasonable people in the society" (2018, 549), p. 549. The dilemma of justifying epistemic and particularly normative assumptions in the context of pluralistic societies may thus be overcome by referring to universal principles that establish the shared standard as a frame of reference.

Although Binns' focus is on a more specific problem, it bears similarities to deliberative order ethics not only in that he draws from ethics and political philosophy, but also in that he uses a similar problem description as starting point. Hence, both approaches start from the question of how conflicts between different values and interests triggered by the use of AI-based systems are to be solved, given the assumption of reasonable value pluralism. The solutions, however, as provided by Binns' approach to algorithmic accountability in terms of public reason on the one hand and our proposal for an approach of deliberative order ethics on the other hand, show two main differences. First, our proposal starts at an intermediate level of business ethics, while Binns like Rahwan (2018) starts at the wider political level of democratic societies. Secondly, Binns' account of public reason provides for value conflicts to be resolved by reference to universal principles, while order ethics attempts to refrain from assuming any universal normative principles. At this point, Binns emphasises that in advocating a public reason-flavoured form of algorithmic

accountability, no particular form of public reason should be presupposed. Instead "the precise content of these common principles is expected to emerge from a process of reflective equilibrium between equal citizens" (Binns 2018, 550), p. 550. To some extent, the process of reflective equilibrium on a societal level to identify universal ethical principles is similar to the deliberative process of order ethics to describe community specific ethical rules. Thus, while Binns seeks to resolve pluralistic conflicts of values in the context of AI by establishing principles and rules at a societal level, we believe that from a business ethics perspective, such rules are most effective and practicable at the community level in terms of specific fields of application. Would it not be plausible to assume that for principles of algorithmic accountability too, different rules might be useful depending on the area of application, but still general enough as it is shared by a group of stakeholders, an economic community? Would it not be possible, for example, to have different shared epistemic and normative standards for algorithmic accountability, depending on whether an online retailer refuses to provide me with a specific offer or whether my application for a vocational training position has been rejected? And still others when I receive my tax return? Despite the differences in the proposed responses, we find in Binns' proposal a corroboration of our concept of deliberative order ethics as it similarly highlights the issue of value conflicts underpinning AI ethics and seeks a solution based on rules and standards developed by public-deliberative dialogue.

Moreover, Wong (2020) introduced a deliberative approach to the question of algorithmic fairness based on the accountability for reasonableness framework (AFR). Similar to the criticism of a too limited technical focus raised in section 2, he argues that algorithmic fairness is mainly conceived as a technical challenge (Wong 2020, 227), p. 227. However, as it turns out, the concept of fairness is in itself controversial and different definitions exist, each with their own implications. Above all, however, this shows that firstly, it is mathematically impossible for an algorithm to fulfil different fairness measures at the same time, i.e. fairness claims other than those implemented are necessarily violated. Secondly, trade-offs between fairness and other factors in the design of algorithms arise, e.g. between fairness and performance or accuracy, or between fairness and safety (Wong 2020, 229), p. 229. As a result, algorithmic fairness is ultimately a question of conflicting values and interests of involved stakeholders and, in this sense, an inherent political task. In response to this challenge, Wong proposes to rely on the accountability for reasonableness framework (AFR) based on Daniels and Sabin (2002). Recognising the pluralistic nature of liberal democracies, the AFR presents a procedural framework to establish a "'process or procedure that most can

accept as fair to those who are affected by such decisions. That fair process then determines for us what counts as fair outcome'" (Daniels and Sabin 2002, 4, quoted from [13]), p. 233. To this end, AFR formulates four conditions in order for decision-making processes to be considered fair and legitimate: a publicity condition, a relevance condition, a revision and appeals condition, and a regulative condition. What Wong thus proposes is to use a procedural account of public deliberation to engage in a "genuine exchange of reasons" and facilitate social learning to find common ground on appropriate fairness measures which are acceptable to all. In this way, Wong's approach is similar to Binns' (2018) proposal in that both emphasise the political dimension of conflicting values and develop a framework for societal responses. Yet while Binns deduces his concept especially against the backdrop of reasonable pluralism, Wong argues on the basis of the internal features of algorithmic fairness (Wong 2020, 239), p. 239. Both Binns and Wong thus explore the question of how a shared normative basis for an ethical approach to AI can be found in the face of differing and conflicting values and interests, once in relation to algorithmic accountability and once in the context of algorithmic fairness. Wong's proposition of an AFR differs from our approach to deliberative order ethics mainly by its explicitly practical claim (Wong 2020, 241), p. 241, which offers concrete criteria for process design in order to determine which values or which conceptualisation of individual values should be adopted in the design and use of AI systems. In this way, Wong's proposal can be seen as a supplement on a practical level of implementation. Regarding the implementation of deliberative or ethics, in particular the revision and appeals condition and criteria for the development of rules require further elaboration. What remains open, however, is how AFR's criteria should be applied in practice. Insofar as the proposal addresses specific AI systems on the one hand, but on the other hand is based on the overarching level of the political system as a whole (deliberative democracy) (Wong 2020, 238), p. 238, the connection between the two levels does not seem trivial. It is here that the added value of our account of deliberative or ethics comes into play as it attempts to provide a more precise description of who (firms) wants to achieve what (establishment of rules) and how (deliberative stakeholder engagement) and at which level (economic communities). This kind of methodological middle step of normative political theory on the one hand, and precise criteria for process design on the other, not only facilitates the practical application but also allows normative theory to be differentiated and expanded over different levels.

Finally, Whittlestone et al. (2019) and Rosenbaum and Fichman (2019) explored the wider societal dimensions of ethical issues in the use of AI. Rosenbaum and Fichman, like Binns (2018), focus on the question of algorithmic accountability and point out the complexities of technical and sociotechnical approaches. Summarising different ways forward they point to the societal and political dimensions of digital justice "by moving away from a focus on the algorithm itself" (Rosenbaum and Fichman 2019, 243), p. 243. As mentioned above, Whittlestone et al. (2019) argue that the normative dimension of AI ethics should be understood as tensions between different values and principles and thus agree with our proposal to take value conflicts as a starting point as well as with the ideas of Rahwan (2018), Binns (2018), Wong (2020). They echo the concern that fundamental normative conflicts may not be resolved with a technical or principled view (B. Mittelstadt 2019; Greene, Hoffmann, and Stark 2019; Green 2020). Rather, they underline the difficulty to achieve acceptable results when dealing with value conflicts or tensions: "Making these trade-off judgements will be a complex political process. Weighing the costs and benefits of different solutions can be an important part of the process but alone is not enough, since it fails to recognise that values are vague and unquantifiable, and that numbers often hide complex value judgements. In addition, resolving trade-offs will require extensive public engagement, to give voice to a wide range of stakeholders and articulate their interests with rigour and respect" (Whittlestone et al. 2019, 199), p. 199. Moreover, trade-offs may not be unavoidable, e.g. if further research and development promises solutions that reduce or even avoid trade-offs (Whittlestone et al. 2019). Should we then wait for the technology to be applied in the future or should we use existing applications? Or does this perhaps point to precisely those instances where a technical, AI-based solution may not be the best alternative (Green 2019)? Whittlestone et al. (2019) therefore stress the need for stakeholder engagement and deliberation in order to evaluate costs and benefits and find agreement on acceptable trade-offs. Our account of deliberative order ethics picks this up and introduces a normative theory of contractualist business ethics to provide a both normatively firm and practicable approach to AI ethics.

Overall, the discussion shows that a broadening of AI ethics towards the fundamental normative conflicts between different values and interests on a societal level is needed (Rahwan 2018; Binns 2018; Wong 2020; Whittlestone et al. 2019; Rosenbaum and Fichman 2019). We understand the proposals by Rahwan (2018), Binns (2018), Wong (2020) and Whittlestone (2019) as an affirmation that the starting point of value conflicts as adopted by deliberative order ethics is suitable for addressing the ethics of AI on a comprehensive societal level. We also agree with Rahwan (2018), Wong

(2020), Whittlestone (2019) and, to some extent, Binns (2018) that in view of the conflicts and trade-offs to be negotiated, a participatory and deliberative approach is appropriate for negotiating a fair distribution of costs and benefits and acceptable outcomes in a mutual exchange between affected stakeholders. Based on contractualist business ethics, our approach further specifies that firms should involve not only stakeholders of a specific business model, but the whole economic community in the sense of all stakeholders involved and potentially affected by the use of AI systems in a specific application area, say recruiting, not least from civil society. Engaging in participatory deliberation, firms should ultimately establish rules that enable mutual advantages to be created for the entire economic community. Thus, while the basic theoretical approach is similar to other current contributions, our approach is distinct in that it introduces a business ethics concept based on a solid theoretical framework. We argue that a business ethics perspective is useful for linking normative-theoretical considerations to the relevant implementation context of AI innovation, to both conceptually enhance AI ethics and improve its practical application. In this respect, our proposal of deliberative order ethics provides a valuable addition to the existing contractualist and deliberative proposals in AI ethics.

Recently, Himmelreich (2019) argued that AI ethics should turn to political philosophy in order to take greater account of the collective decisions that are evoked by the use of AI systems. Political philosophy would then be able to add three basic concerns to the conceptual toolkit of AI ethics: reasonable pluralism, individual agency, legitimate authority. As explained above, contractualist business ethics is fundamentally based on the question of preserving reasonable pluralism and the possibilities of legitimate authority. Along these lines, our theory-based proposal contributes to complement AI ethics with political philosophy. Insofar as we are developing a concept of business ethics on such a theoretical basis, we even go one step further and argue that AI ethics needs not only political philosophy but also normatively sound contributions of business ethics (Luetge, Armbrüster, and Müller 2016; Heath, Moriarty, and Norman 2010; Moriarty 2005). It is precisely from such an integrated approach that AI ethics can be pushed forward both in terms of normative concepts and in practice.

4.2.8 Problems solved? Addressing current shortcomings with deliberative order ethics

Recognising the beginnings of a third wave, we started our reasoning with a critical discussion of the current state of AI ethics. To this end, we summarised five shortcomings of a first and second wave of AI ethics. In this section we will therefore briefly discuss whether or not and to what extent the concept of deliberative order ethics may successfully address current weaknesses.

1. AI ethics neglects the business context of developing and employing AI systems

The first shortcoming concerned the focus on the technical level of AI systems, which tends to neglect the integration in an entrepreneurial and wider societal context. In response to this, deliberative order ethics offers a valuable contribution as it starts at the level of firms and their business practices which may create value conflicts. The deliberative approach involving the stakeholders of an economic community does not focus on the ethical design of the technical system alone, but rather on the question of how the use of AI in a specific societal context is acceptable for all involved, i.e. how it can be implemented to the benefit of all. The object of ethical scrutiny is thus not only the AI system on a micro level but the AI system in the context of its commercialisation, and its impact on markets, the environment, individuals and society. In this way, primarily responsible actors are not individual developers but firms on an organisational level and their strategic behaviour in economic communities. By introducing a business ethics perspective, the business context of the use of AI systems is systematically taken into account. At the same time, the political dimension of AI ethics is reflected by focusing on value conflicts and deliberatively negotiating acceptable trade-offs with all stakeholders, not least from civil society. Drawing from political philosophy, deliberative order ethics thereby is able to "address fundamental normative and political tensions" (B. Mittelstadt 2019), p. 501, prompted by AI systems.

2. AI ethics is biased toward a technological solutionism

Closely connected to the first issue of too narrow a focus on technical systems alone, the second problem refers to the problem of technological solutionism. By focusing on the ethical design of AI systems, the question of whether or not a particular AI system should be built and employed in the first place or whether there is perhaps another, possibly non-technical alternative that can better solve the problem at hand moves out of sight. Here, too, deliberative order ethics seems to offer an

approximation to the problem by broadening the scope of AI ethics. Insofar as rules for a fair distribution of costs and benefits and acceptable trade-offs are negotiated in a deliberative procedure, what is at stake is a more comprehensive evaluation of AI-based business practices. This includes the question of whether the use of AI systems appears appropriate and reasonable in view of the identified costs and benefits. Although its goal is not to devise possible alternatives, the deliberative establishment of rules defines the framework (level playing field) within which the use of AI systems is socially acceptable. By formulating the conditions accordingly, minimum thresholds can be set and certain AI applications may thus be ruled out.

3. AI ethics succumbs to an individualist focus

Thirdly, we pointed out that the focus on AI systems and their ethical design leads to ethical action primarily to be located at the level of individuals such as data scientists and developers. However, this fails to recognise the role of the organisational level, both internally in terms of governance mechanisms and corporate culture and externally in terms of its integration in markets and the wider societal contexts. Here the perspective from business ethics helps to bring the role of firms as organisations into focus. In doing so, deliberative order ethics concentrates less on internal aspects than on external interactions and the behaviour of businesses in society. Accordingly, responsibility for ethical behaviour is no longer attributed (only) to developers but to firms as social actors who need to establish general rules for the ethical use of AI systems on the basis of their own interests. In this sense, firms have the responsibility to initiate or actively participate in respective deliberative processes if they intend to establish a new business segment or expand an existing one by using AI.

4. AI ethics is problematic in its implementation and lacks accountability and clear impact

The fourth weakness relates to the difficulty of making abstract principles manageable in practice, identifying effective approaches and ensuring their normative soundness (gap between a variety of tools and normative justification). Here the approach of deliberative order ethics can only partly provide an adequate answer as it fell outside the scope of this article to spell out its practical application. Nevertheless, in response to the observed implementation problems, our approach advocates a systematic combination of normative theory with an applied perspective of business ethics. However, the concept needs to be fleshed out for its implementation and its effectiveness needs to be critically evaluated.

5. AI ethics does not clarify its link to legal regulation

Finally, we pointed out the ambiguous relationship between AI ethics and legal regulation, which has been criticised by a number of authors. Here, our proposal provides a thorough clarification. Based on a contractualist theory of normative business ethics (Luetge 2013) and the concept of incomplete contracts (Hart 2017), the complementary character of ethics can be specified (cf. Luciano Floridi 2019). Deliberative order ethics highlights the relevance and need of legislation and hard law regulating the development and use of AI in order to set clear and binding rules enabling fair competition. It is only in addition to these regulations that order ethics seeks to address inevitable gaps for ethical rules. Thereby, ethics may also function as participatory creation and testing of ethical rules for as long as the legislative process is still under way (cf. Larsson 2020), and rules might even become legislation at some point. According to the complementary nature of the relationship, deliberate order ethics should thus under no circumstances provide a basis for avoiding or delaying legislation.

To conclude, complementing AI ethics with deliberative order ethics thus offers several benefits. By integrating ethical considerations on the business practices surrounding the use of AI systems in society, it forces firms to identify and analyse the diverse impacts the employment of AI may have, to discuss costs and benefits from a comprehensive perspective with all stakeholders from the community and explicitly formulate rules for acceptable trade-offs that allow added value for all. Thereby value conflicts caused by critical business practices become the subject of an open dialogue. As a result, both new and already existing problematic business practices can be revealed and put up for discussion (cf. Binns 2018). Furthermore, the deliberative approach not only helps uncover relevant values and interests in society, it also initiates the weighing up and explicit balance of different societal values, from economic to social and ecological values. Establishing societal standards for a desirable use of AI, deliberative order ethics offers an avenue designed to make the use of AI beneficial for everyone in society. Beyond avoiding adversarial effects, this makes pluralistic value creation the ultimate ambition of AI ethics.

4.2.9 Conclusion

The increasing use of AI systems not only presents great opportunities for many important areas of society such as medicine or climate protection, it also raises profound ethical questions and challenges fundamental societal values. Recognising these impacts, the field of AI ethics emerged developing both theoretical guidelines as well as practical tools addressing issues such as unfair discrimination or algorithmic accountability.

In this article, we introduced a procedural account of a deliberative order ethics to complement AI ethics. To this end, we first presented the current state of AI ethics which results in our review in two first waves of AI ethics. At the threshold of the beginning of a third wave we consolidate different concerns by arguing that in its current form AI ethics is facing at least five crucial shortcomings: AI ethics tends to neglect the business context of developing and employing AI systems, it is biased toward a technological solutionism, succumbs to an individualist focus, is problematic in its implementation and lacks accountability and clear impact, and does not clarify its link to legal regulation. Building on this critique, we first introduced the contractualist concept of normative business ethics called order ethics. Contrasted with ISCT as the most proliferated theory of contractualist business ethics, we argue that deliberative order ethics provides an adequate approach to deal with the complex value conflicts firms may trigger through AI in pluralistic societies. Order ethics holds that these conflicts should be resolved by adequate rules so that ethical behaviour does not lead to a competitive disadvantage to which stakeholders agree based on mutual benefits. Secondly, we proposed a procedural expansion arguing that it is through participation and deliberation that stakeholders of an economic community may adequately discuss costs and benefits and agree on rules for acceptable trade-offs when using AI systems in their respective field. Thereby, deliberative order ethics ultimately seeks to make the use of AI systems a matter of pluralistic value creation. The role of ethics in AI thus becomes, among other things, to ensure that AI creates diverse societal and ecologic value in combination with financial business value.

By complementing AI ethics with an approach of business ethics we aim to integrate the level of business practices into the considerations of AI ethics and highlight the organisational role of firms for achieving ethical AI. Since AI systems and most other emerging technologies are at least commercialised and brought to society by firms through new business models or enhanced products and services, we believe that one cannot achieve truly ethical AI without addressing key issues of business ethics. From a favourable point of view, building "ethical" AI systems that are then part of

questionable and dodgy business practices and markets does not seem to cover the whole picture. At worst, it is part of a deceitful strategy and irresponsible. Some perspectives in AI ethics seem to implicitly assume an opposing relation of business and ethics which ultimately leads to the conclusion that truly ethical AI can only be possible beyond businesses and markets in their current logic and structure. This, however, means admitting that, at least in the short and medium term, ethical AI business practices cannot gain wider application. Although we are somewhat sympathetic to such a sincerely critical and more idealistic view, our ambition is to strive for realistic change to achieve ethical AI at all, given the current circumstances. At the critical point at which AI innovation currently stands, it is thus particularly important to reflect the normative foundations of both economics and businesses and to make AI ethics an endeavour of business ethics too. To this end, we have made a first contribution with this article.

Combining two previously separated fields of research, this article shows some important limitations. First, as we focused on demonstrating how a business ethics perspective may provide a valuable complementation to AI ethics addressing some of its current shortcomings, a detailed description of the concept of deliberative order ethics and what a concrete application might involve exceeded the scope of this article. Among other aspects, this would entail more conceptual detail with regard to criteria for deliberation processes, rules and consent as well as an exemplary description of its implementation in practice. Secondly, our analysis of the first two waves of AI ethics only covers major trends and was thus unable to do full justice to the diverse and dynamic field that is emerging today. An adequate review for the purpose of a systematic critique of the field would provide the subject of an entire article. Alone the focus of this article was a different one.

These limitations, however, also provide the basis for further research. Future research agendas should seek to both systematise and consolidate the very diverse and dynamic field of AI ethics as well as foster the development of solid practical approaches. More specifically, further research needs to develop practical approaches and tools based on well-founded normative claims and evaluate their effectiveness empirically.

Finally, this article contributed to a diversification of causation narratives about (un)ethical AI (Greene, Hoffmann, and Stark 2019): without taking into account the role of business practices and markets, AI ethics risks never reaching its goals. Therefore, political philosophy and business ethics are urgently needed complements to ensure that AI ethics remains a theoretically sound and practically effective effort.

4.3 Social acceptance of green hydrogen: Building trust through participation

4.3.1 Publication details

The third co-authored paper by Häußermann, J.J.; Kirsch, T.-C.; Maier, J. M.; Kaiser, S. and M. Schraudner is submitted and currently under review at *Energy Policy* (ISSN: 0301-4215). Author contributions are: research design: J.H., S.K., M.S.; qualitative data collection and analysis: J.H., T.K.; quantitative data collection and analysis: J.H., M.M., T.K.; wrote the paper: J.H.

4.3.2 Abstract

Social acceptance presents a major challenge for Germany's transition to green energy. As a power-to-x technology, green hydrogen is set to become a key component of a future sustainable energy system. With a view to averting conflicts like those surrounding wind energy, we have investigated social acceptance of green hydrogen at an early stage in its implementation, before wider rollout. Our study uses a mixed-method approach, wherein semi-structured interviews (n=24) and two focus group workshops (n=51) in a selected region in central Germany serve alongside a representative survey (n=2054) as the basis for both understanding social attitudes and reaching generalisable conclusions. Overall, it is possible to observe both a marked lack of knowledge and a large degree of openness towards green hydrogen and its local use, along with high expectations regarding environmental and climate protection. We reach three key conclusions. Firstly, acceptance of green hydrogen relies on trust in science, government, the media, and institutions that uphold distributive justice, with consideration for regional values playing a vital role in establishing said trust. Secondly, methodologically sound participatory processes can promote acceptance, and active support in particular. Thirdly, recurrent positive participatory experiences can effectively foster trust. Accordingly, we argue that trust should be strengthened on a structural level, and that green hydrogen acceptance should be understood as a matter of responsible innovation. As the first empirical investigation into social acceptance of green hydrogen, and by conceptually interlinking acceptance research and responsible innovation, this study constitutes an important contribution to existing research.

4.3.3 Introduction

In order to achieve greenhouse gas neutrality by 2050 and honour the Paris Agreement as well as to implement the sustainable development goals (SDGs), Germany needs renewable energy sources to form the cornerstone of an environmentally friendly and sustainable energy supply system (Bundesregierung der Bundesrepublik Deutschland 2010). The government's „National Hydrogen Strategy" (Bundesregierung der Bundesrepublik Deutschland 2020) accordingly identifies green hydrogen, which is produced via electrolysis powered by renewable electricity, as a key component of such a system. On a European level too (European Commission 2020), the potential of green hydrogen to play a key role in the European Green Deal for the European Union (European Commission 2019) has been recognized (Brudermüller et al. 2021). However, development and above all implementation of this technology are still at an early stage. The resulting lack of practical experience with it, especially among the general public, makes studying social acceptance of green hydrogen challenging (Hildebrand, Gebauer, and Taubitz 2019). Yet therein also lies a valuable opportunity: the earlier relevant acceptance factors can be identified, the more effectively long-term, sustainable solutions can be developed, ensuring wide social adoption and use across various industries.

As initial explorations of power-to-x technologies emphasise, participation has an important role to play in this regard (Hildebrand, Gebauer, and Taubitz 2019; Ashworth et al. 2019). In the debate surrounding renewable energy infrastructures, this is by no means a novel insight: participation has been considered key to Germany's transition to green energy for a number of years (cf. Holstenkamp and Radtke 2018). Creating opportunities for public involvement can help resolve or mitigate potential conflicts arising from the creation of renewable energy facilities, for instance when selecting locations for wind farms. Moreover, openness and transparency increase the legitimacy of planning and decision-making processes – regardless of their actual outcomes. Social acceptance is thus not solely dependent on the properties of a given technology, but is likewise influenced by interaction and dialogue among various stakeholders.

This study therefore focusses on the role of participation in green hydrogen acceptance. The aim is to show how, as a matter of responsible innovation, participation can promote trust and lay a sustainable foundation for social acceptance of green hydrogen.

To this end, we chose a mixed-method approach, combining a representative survey of the German population (n=2054) – designed to both gauge green hydrogen acceptance and test six hypotheses on the relationship between acceptance, trust and participation – with findings derived from qualitative interviews (n=24) and two focus group workshops (n=51) conducted as part of a case study carried out in and around Leipzig (McCusker and Gunaydin 2015). The survey results were subjected to quantitative analysis for the sake of drawing reliable and representative conclusions, while also representatively validating the qualitative findings from our case study. Conversely, our case study findings facilitate a better understanding of the relationships confirmed by our quantitative survey.

This paper is structured as follows. In the next section [4.3.4], we define our research question against the background of the current state of research on the connection between participation and social acceptance of sustainable energy system transformation. After that, we present our quantitative [4.3.5] and qualitative [4.3.6] findings, which are then discussed in terms of the relationship between participation, trust and green hydrogen acceptance [4.3.7]. Finally, we point out important implications and limitations of the study, and suggest avenues for further research [4.3.8].

4.3.4 Background

Green hydrogen has the potential to become a cornerstone of the global transition to green energy (Dincer 2012; Hosseini and Wahid 2016; van Renssen 2020). Firstly, converting renewable electricity to green hydrogen makes it possible to store weather-dependent regenerative energy derived from wind and sunlight (Ogden 1999; Ozarslan 2012; Zhang et al. 2016). Secondly, hydrogen-powered fuel cells and hydrogen-based synthetic fuel constitute possible alternatives to direct electricity usage and battery-powered technologies in transport (Farrell, Keith, and Corbett 2003; Singh et al. 2015) and heating (Dodds et al. 2015). Furthermore, in areas such as the steel and chemical industry, green hydrogen technology provides the only route to decarbonisation (Bhaskar, Assadi, and Nikpey Somehsaraei 2020; Otto et al. 2017). At the same time, green hydrogen's potential role in Germany's energy transition is a matter of not only technical feasibility, but also socially accepted implementation (A. Schmidt, Canzler, and Epp 2019). This becomes all the more important as the resource-intensive construction of infrastructure requires decisions with long-term impact, although some existing infrastructures, such as natural gas pipelines, may also be used for hydrogen.

4.3.4.1 Social acceptance of renewable energy innovation and the role of participation

Although the majority of Germany's population has a fundamentally positive attitude towards the country's energy transition, conflicts regularly arise during the planning and implementation phases of specific projects. The „social gap" (Bell et al. 2013; Bell, Gray, and Haggett 2005) between strong approval voiced in opinion polls on the one hand, and local protests on the other, shows that social acceptance takes place on different levels (Wüstenhagen, Wolsink, and Bürer 2007). Public reaction to new energy environments is thus not just predicated on aspects of the technologies themselves, such as potential safety issues. Local context and the ways in which people are personally affected likewise influence their perception and assessment of changes to their everyday surroundings (Perlaviciute and Steg 2014; Upham, Oltra, and Boso 2015).

In the past, public rejection was often seen as symptomatic of the NIMBY ("not in my backyard") syndrome, yet this simplified view is now considered outdated (Devine-Wright 2009, 2005; Wolsink 2000). Current approaches suggest that local changes can just as easily elicit positive reactions „in situations of good 'fit' between symbolic meanings associated with both place and project" (Devine-Wright 2011, 341), with

local context and the community's day-to-day life providing possible symbolic, emotional and ideational points of reference (Batel 2018; Bosch and Schmidt 2020). This also pertains to the relationships between the stakeholders of a given project. In particular, the presence or absence of a sense of trust and fairness among stakeholders helps determine the public's reaction to new energy technologies (Huijts, Molin, and Steg 2012). Accordingly, in both research and practice, the question of how increased participation can promote renewable energy acceptance has become increasingly relevant (Hildebrand, Rau, and Schweizer-Ries 2018; Holstenkamp and Radtke 2018).

In contrast to renewable energy technologies such as wind and solar power, green hydrogen is still largely unknown to the majority of the German population (Hyacinth Project 2017; Zimmer and Welke 2012). To date, only a small number of public participation projects and focus group studies have examined the German public's attitude towards hydrogen technologies in general. In addition to sparse knowledge and a lack of direct experience, these studies reveal a fundamental openness that could turn into either approval or rejection, depending on the information made available and the dynamics of communication (Scheidler and Pfaff 2019) – as well as doubts as to whether hydrogen technologies are being developed and used to serve the common good, and not just business interests (Zimmer and Welke 2012; Hytrust 2013). Research carried out in the United Kingdom and the Netherlands reveal similar findings, identifying distrust of industry and public institutions as a significant barrier to acceptance (Huijts, Molin, and van Wee 2014; Ricci, Bellaby, and Flynn 2010; Mumford and Gray 2010). Against this background, we maintain that social acceptance of green hydrogen should be investigated at a very early stage in its implementation, with a particular focus on the role of trust.

4.3.4.2 Linking social acceptance with the concept of responsible innovation

The concept of responsible innovation centres on the development of new technologies in harmony with social values (Stilgoe, Owen, and Macnaghten 2013; René Von Schomberg 2013; Rip 2014; Ruggiu 2015; Burget, Bardone, and Pedaste 2017). The corresponding academic field comprises not only conceptual and empirical research, but also the development of practical approaches, including methods and tools that help shape processes, organisations or the innovation system as a whole (Long, Iñigo, and Blok 2020; Lubberink et al. 2017). Responsible innovation aims not just to avoid ethically, socially and ecologically undesirable

consequences, but also to address specific societal challenges through innovations that contribute to the common good. Operating within a participatory governance framework, innovation processes should accordingly involve not only directly responsible organisations, but also more broadly affected societal actors, society at large, or individual members of the public (Richard Owen et al. 2013a). The underlying premise is that early integration of a variety of stakeholders makes it possible to incorporate multiple perspectives – and thus also social values – into the development of new technologies. In a process thus marked by inclusion and responsiveness, existing dynamics between actors within the innovation system can change and new relationships be forged, while a greater degree of democratisation helps establish mutual trust among civil society, science, business and government (Stilgoe, Owen, and Macnaghten 2013; V. Blok, Hoffmans, and Wubben 2015).

Despite being concerned with the development and implementation of new technologies in line with social values and for the benefit of society – a process in which participation plays a central role – the concept of responsible innovation has, to date, not been systematically linked with acceptance research. Within acceptance research, on the other hand, explicit connections have been drawn to responsible innovation, most prominently by Pellizzone et al. (2017, 2015). Were the two fields systematically interlinked, it could help uncover their respective blind spots and open the door to mutual learning and synergy. That this has not yet happened is all the more surprising given that the concept of responsible innovation has been incorporated into research on the global transition to green energy (Koirala, van Oost, and van der Windt 2018; Correljé et al. 2015), and considering that acceptance has always been considered a (side-)effect or even one of the aims of responsible innovation (René Von Schomberg 2012; Richard Owen et al. 2013a).

We suspect that there are two reasons for this. Firstly, the focus of acceptance research falls primarily on individuals, groups, communities or entire societies as (potential) end users or as (potentially) affected by a given technology. The technology itself, however, is viewed from a passive, external perspective as an object of acceptance that is essentially taken as given (Upham, Oltra, and Boso 2015). Responsible innovation, on the other hand, focusses on the circumstances and process surrounding the development of the technology in question, with the aim of substantially influencing said process and altering its outcome in response to societal needs. Secondly, responsible innovation aims to effect systemic change, its medium- to long-term goal being the democratisation of innovation via sustainably changed relationships between different actors (Richard Owen et al. 2013a; Carayannis and

Campbell 2012b; Braun and Griessler 2018). By contrast, acceptance research mainly focusses on specific use cases, the end goal being the successful introduction and sustainable local or general adoption – i.e. acceptance – of a given technology.

This paper constitutes an argument for more decisively and systematically interlinking these two fields of research and practice. Using the example of green hydrogen, we aim to show how this can benefit the investigation and promotion of social acceptance from the perspective of responsible innovation.

4.3.4.3 Research question

In this paper, we investigate social acceptance of green hydrogen in Germany, with a particular focus on the role of participation. Working from the assumption that trust is crucial for social acceptance, and that participation should therefore aim to promote trust, we ask the following research question: can participation promote social acceptance of green hydrogen – and if so, how?

4.3.5 Quantitative data

4.3.5.1 Methods

A representative survey was conducted between 15 and 25 May 2020, in the form of an online questionnaire completed by a total of 2054 respondents. The results of the survey were weighted and are representative of the German population aged 18 and above (fig. 1). The standardised German-language questionnaire was developed by the authors of this paper, and the fieldwork carried out by an external contractor.

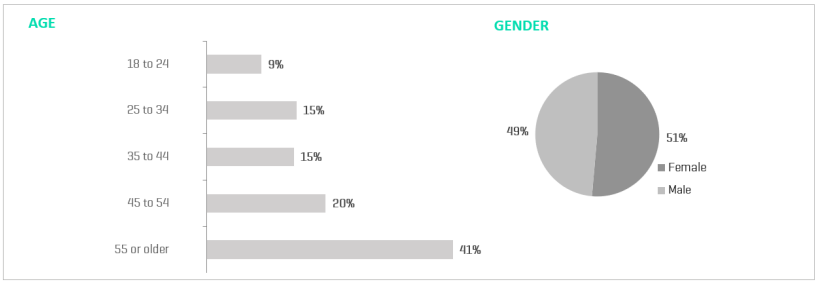

Fig. 1: Structure of the quantitative data sample

4.3.5.1.1 Description of questionnaire and variables

The questionnaire consisted of 97 questions grouped into 27 sets. In addition to general demographic questions regarding age, place of residence, marital status, income, household size and political orientation, respondents were given questions related to green hydrogen acceptance. Environmental awareness was assessed via 23 questions and – as in the federal government's most recent study on environmental awareness in Germany (Bundesministerium für Umwelt 2019) – mean values were subsequently calculated to respectively represent cognitive environmental awareness, emotional environmental awareness, and environmentally conscious behaviour on a scale of zero to ten. Further questions focussed on experiences with participatory processes and the perceived competence and trustworthiness of a variety of actors from the fields of government (comprising both local and national government, as well as European institutions), the media (comprising print media as well as public and private broadcasting), research and education (comprising research facilities and academic institutions), business (comprising small and large-scale enterprises) and respondents' own private sphere. Additional factors assessed via the questionnaire include respondents' perceived self-efficacy, their willingness to become involved in promoting or opposing renewable energy use in their own city or municipality, their

familiarity with hydrogen in general and green hydrogen in particular, what impact they expect green hydrogen to have on areas such as safety, mobility and environmental protection, and the importance they attach to public involvement in green hydrogen adoption. The scales used are explained in more detail in section 3.2 below.

4.3.5.1.2 Statistical analysis

In line with our research question, the following statistical methods were employed in analysing the survey data. In an initial exploratory analysis, logistic regression was used to determine the factors that influence local approval of green hydrogen. A second logistic regression analysis uncovered factors influencing the public's willingness for active involvement in this regard. Logistic regression was chosen as analytic method given the absence of a normal distribution of individual variables, and was performed using binary dependent variables. Effect size was calculated following Cohen (1992). Due to its exploratory nature, this initial stage of the analysis encompassed numerous independent variables. Given the variables' level of measurement and partial lack of normal distribution, a Spearman rank correlation was subsequently carried out (Hollander, A. Wolfe, and Chicken 2015). Differences between groups were then analysed using the nonparametric Wilcoxon test, since the variables did not satisfy the assumption of normality (Gehan 1965). In the case of multiple tests, Bonferroni correction was used to counteract alpha error accumulation (Bland and Altman 1995).

4.3.5.2 Results

4.3.5.2.1 Descriptive findings on green hydrogen acceptance in Germany

While 85% of respondents indicate having heard of hydrogen, the same is true for only 26% in the case of green hydrogen (fig. 2). Here, significant differences can be observed between age groups: older individuals are more familiar with hydrogen as such, whereas younger people are more likely to have heard of green hydrogen (fig. 3). Familiarity with (green) hydrogen is also seen to differ significantly in accordance with level of education, with those in possession of or currently studying towards a university degree or equivalent qualification being more likely to have some knowledge of it (fig. 4). No significant regional differences can be observed in this respect.

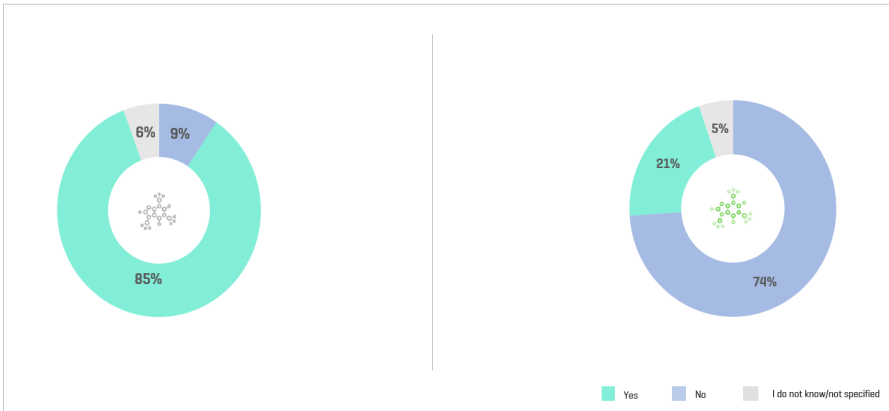

Fig. 2a: Familiarity with hydrogen

Fig. 2b: Familiarity with green hydrogen

Fig. 3: Familiarity with hydrogen and green hydrogen by age

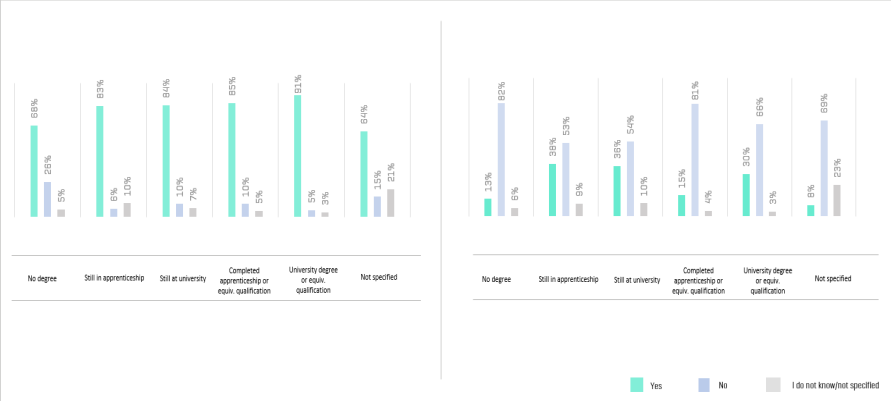

Fig. 4a: Familiarity with hydrogen by level of education

Fig. 4b: Familiarity with green hydrogen by level of education

Among respondents, the most well-known use of hydrogen is within the field of mobility (as indicated by 70%), followed by the energy supply sector (56%) and industrial settings (48%). Only 37% of respondents feel able to form an opinion of

hydrogen technologies, in contrast to other energy technologies such as solar (62%) and wind power (60%). 64% of respondents expect green hydrogen to have a positive impact on environmental protection and sustainability, followed by mobility and infrastructure (52%), regional value creation (43%), and job creation and structural change (40%). Only 7% of respondents expect green hydrogen to have a negative impact on public and personal safety, whereas 60% foresee no or even a positive impact, and 33% declined to comment. Here it is worth noting that 87% of undecided respondents, along with 61% of those who foresee a negative impact on safety, nonetheless have a very or somewhat positive attitude towards local green hydrogen use.

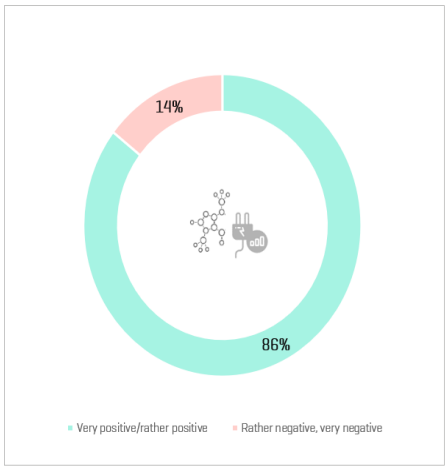

Fig. 5: Local approval of green hydrogen in Germany

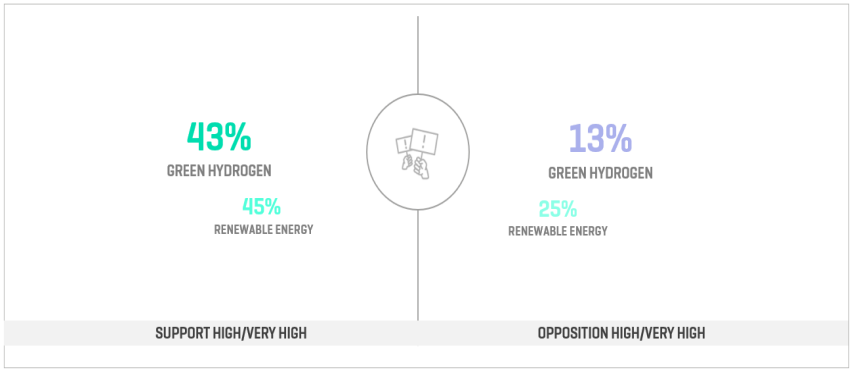

Fig. 6: Local active support for and active opposition to green hydrogen and renewable energies in general in Germany

86% of respondents indicate feeling very or somewhat positive about green hydrogen usage in their own cities or municipalities (fig. 5). Whereas 43% of those in favour are very or somewhat willing to become actively involved in promoting local adoption of green hydrogen, only 13% of those opposed indicate a willingness for active involvement (fig. 6). Furthermore, knowledge is seen to have a highly significant influence on approval – 92% of respondents familiar with green hydrogen approve of

its use in their local environment, as opposed to 85% of those unfamiliar with it – and an even clearer impact on potential active support: only 44% of those without knowledge indicate a willingness to become actively involved, in contrast to 69% of those with knowledge of green hydrogen.

The overall high level of approval despite limited familiarity with green hydrogen reflect the findings of Achtenberg et al. (2010) on the role of knowledge – in that it only leads to increased acceptance under certain circumstances and in conjunction with factors such as trust and cultural predispositions. The substantial level of local approval shows that local acceptance is better explained in terms of place attachment than as a matter of NIMBY reactions (Devine-Wright 2009, 2005; Devine-Wright and Howes 2010). Overall, the combination of limited knowledge and fundamental openness towards hydrogen technologies revealed by the survey is consistent with earlier findings from Germany (Scheidler and Pfaff 2019; Zimmer and Welke 2012), the Netherlands (Achterberg et al. 2010; Huijts, Molin, and van Wee 2014), the United Kingdom (Flynn, Ricci, and Bellaby 2013; Ricci, Bellaby, and Flynn 2008) and Spain (Iribarren et al. 2016). Our findings on the influence of age and level of education on hydrogen acceptance likewise confirm the results of earlier studies (Achterberg et al. 2010; Altmann and Graesel 1998; Zachariah-Wolff and Hemmes 2006). A clear difference can however be observed between public familiarity with hydrogen in general, and green hydrogen in particular. Whereas many people have a rough understanding of the former, having learnt about hydrogen in school or having encountered it in a range of practical settings[12], the term "green hydrogen" has not yet entered the broader public lexicon. Nevertheless, the population is very open to and willing to use green hydrogen, in large measure thanks to its expected positive impact on environmental protection and the achievement of climate targets (A. Schmidt, Canzler, and Epp 2019; Zimmer and Welke 2012). The issue of safety, on the other hand – to which technology experts often attach great importance – turns out not to be a particularly decisive factor for green hydrogen acceptance (Flynn, Ricci, and Bellaby 2013; Ricci, Bellaby, and Flynn 2008).

[12] This is suggested by our qualitative findings from the interviews and focus groups.

4.3.5.2.2 Environmental awareness in Germany and its impact on green hydrogen acceptance

As part of the survey we assessed the environmental awareness in Germany, as conceptualised by Scholl et al. (2015). Compared to the most recent biennial representative survey of the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) (2019) our findings reveal a marked decrease in cognitive environmental awareness (from 7,9 to 4,2), emotional environmental awareness (from 7,2 to 4,8) and environmentally conscious behaviour (from 4,6 to 3,7) (fig. 7).

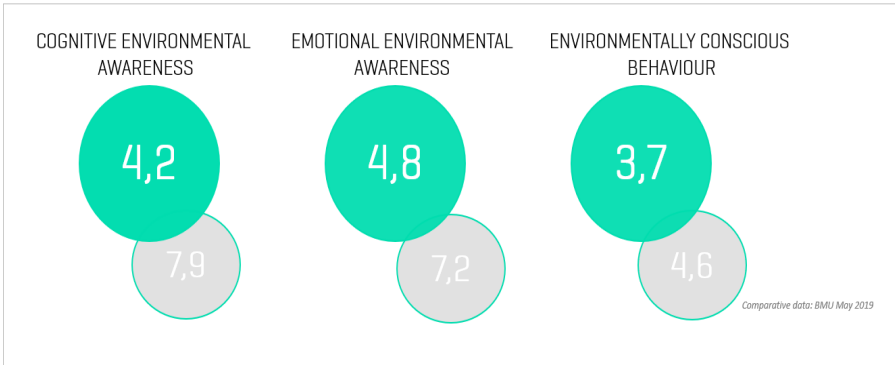

Fig. 7: Level of environmental consciousness in Germany

The significant decrease in environmental awareness can be explained by the drastic shift in public attention that occurred between the two surveys as a result of the COVID-19 pandemic. Whereas environmental and climate protection continually stood at the centre of public discourse in 2018 and ultimately led to the start of the Fridays for Future movement, our survey coincided with the outbreak of COVID-19 in Germany, the effects of which had a particularly strong grip on the public's attention in May 2020. Accordingly, issues related to environmental and climate protection are seen to have lost much of their cognitive and emotional significance. The somewhat less pronounced decline in environmentally conscious behaviour can be accounted for by the fact that patterns of behaviour are slower to change and less easily influenced than cognition and affect. These findings demonstrate that environmental awareness is not yet embedded deeply enough in German society to withstand dramatic shifts in public attention. Moreover, our findings reveal the scope of the challenge involved in restoring – and expanding on – pre-pandemic levels of environmental awareness, and bringing related issues, targets and tasks back to the forefront.

Using two logistic regression analyses, we were furthermore able to determine that cognitive and emotional environmental awareness influence green hydrogen acceptance in different ways. For the first regression model, we assumed that higher levels of cognitive environmental awareness have a positive impact on approval of green hydrogen. Our analysis shows both the model as a whole (chi-squared(23)=236.28, p=.000, n=2054) and the coefficient of individual variables to be significant. When cognitive environmental awareness increases by one unit, local approval increases by 92.3%. The Nagelkere R-squared value is .214, corresponding to a large effect following Cohen (1992).

Our second model rested on the hypothesis that higher levels of emotional environmental awareness positively influence people's willingness to actively promote green hydrogen use. Here, too, a logistic regression analysis reveals both the model as a whole (chi-squared(27)=252.76, p=.000, n=1397) and the coefficient of individual variables to be significant. When emotional environmental awareness increases by one unit, willingness for active involvement increases by 74.9%. The Nagelkere R-squared value is .221, corresponding to a large effect following Cohen (1992).

The correlation observed in both instances firstly underlines the importance of environmental awareness in promoting green hydrogen acceptance (Emmerich et al. 2020; Hyacinth Project 2017). Secondly, it becomes clear that specific measures are needed to promote acceptance as a matter of public attitude on the one hand, and to increase support in terms of active involvement on the other. Whereas the former requires focussing on information and convincing factual arguments, the latter calls for measures that operate on an emotional level. In light of the current urgent need to activate advocates of green hydrogen (Hildebrand, Rau, and Schweizer-Ries 2018; Local Energy Consulting 2020), our findings thus provide a blueprint for the design of measures that effectively promote acceptance. Merely relying on arguments based on technological advantages or logical necessity will not suffice; rather, green hydrogen and its specific practical applications need to be embedded in visions that are as captivating as they are robust.

4.3.5.2.3 Participation, trust and green hydrogen acceptance, part I

Our quantitative survey also served to test six hypotheses on the significance of trust and participation to green hydrogen acceptance.

Building on earlier findings on the role of trust in various actors (Huijts, Molin, and Steg 2012; Midden and Huijts 2009; Montijn-Dorgelo and Midden 2008; Pellizzone et al. 2015; Pellizzone, Allansdottir, and Manzella 2019) and in institutions that ensure the just distribution of costs and benefits (Gross 2007; Huijts, Molin, and Steg 2012; Scherhaufer et al. 2017; Wolsink 2007; Wolsink and Breukers 2010), our first two hypotheses operationalise the relationship between trust and acceptance:

Hypothesis 1.1: *A positive correlation exists between green hydrogen acceptance and the perceived trustworthiness of government, business, science and the media.*
Hypothesis 1.2: *A positive correlation exists between green hydrogen acceptance and high levels of confidence in the fair distribution of the costs and benefits of green hydrogen use.*

Regarding hypothesis 1.1, our findings reveal a significant correlation between local green hydrogen acceptance and the perceived trustworthiness of the scientific sector ($r_s = .319$, $p<.001$, $n = 2088$[13]), corresponding to a moderate effect following Cohen (1992). Local green hydrogen acceptance is likewise seen to correlate significantly with the perceived trustworthiness of government actors ($r_s = .209$, $p<.001$, $n = 2088$) and the media ($r_s = .212$, $p<.001$, $n = 2088$), in both instances corresponding to a small effect following Cohen (1992). The perceived trustworthiness of the business sector, on the other hand, does not correlate significantly with local green hydrogen acceptance. With respect to hypothesis 1.2, our findings show local green hydrogen acceptance to correlate significantly with confidence in the fair distribution of the costs and benefits of green hydrogen use ($r_s = .258$, $p<.001$, $n = 2088$), corresponding to a small effect following Cohen (1992). These findings reveal trust and perceived distributive justice to be relevant acceptance factors for green hydrogen, with trust in the scientific sector emerging as particularly influential.

The next two hypotheses deal with the effect of participation on green hydrogen acceptance, building on earlier findings that reveal participation to have a positive impact on the acceptance of other renewable energy technologies (Baxter 2017;

[13] Summarised values for the perceived trustworthiness of government, the media, science and business are as follows: science: 52% very/somewhat trustworthy – 37% both in equal measure – 11% very/completely untrustworthy; government: 31% very/somewhat trustworthy – 44% both in equal measure – 25% very/completely untrustworthy; media: 25% very/somewhat trustworthy – 31% both in equal measure – 33% very/completely untrustworthy; business: 22% very/somewhat trustworthy – 44% both in equal measure – 34% very/completely untrustworthy.

Langer et al. 2018; Langer, Decker, and Menrad 2017; Lienhoop 2018; Schenk, Hunziker, and Kienast 2007; Suškevičs et al. 2019; Zoellner, Schweizer-Ries, and Wemheuer 2008):

Hypothesis 2.1: *People with participatory experience display higher levels of approval of green hydrogen.*

Hypothesis 2.2: *People who assess their own participatory experiences favourably display greater willingness to actively support green hydrogen adoption.*

With respect to hypothesis 2.1, our statistical analysis reveals significantly higher levels of approval of green hydrogen among people with participatory experience (mean 1,69, mean rank 874,73), compared to those without (mean 1,94, mean rank 1089,70).[14] The asymptotic Wilcoxon test yields a z value of -7,494, a p value of <.001, and an n value of 2088. The r value of .164 corresponds to a small effect following Cohen (1992). Regarding hypothesis 2.2, people who assess their participatory experiences favourably display significantly greater willingness for active involvement (mean 1,88, mean rank 147,49) than those who do not (mean 2,31, mean rank 192,18).[15] The results of the asymptotic Wilcoxon test are z = 3,445, p<.001 and n = 365. Here, the r value is .180, corresponding to a small effect following Cohen (1992). These findings indicate a notable correlation between participation and approval of green hydrogen, and between the quality of people's participatory experiences and their willingness to become actively involved in promoting its use.

The final two hypotheses focus on the relationship between participation and trust:

Hypothesis 3.1: *Positive participatory experiences correlate positively with trust in government, business, science and the media.*

Hypothesis 3.2*: Positive participatory experiences correlate positively with expectations that the costs and benefits of green hydrogen use will be distributed fairly.*

Regarding hypothesis 3.1, our findings reveal a highly significant correlation between positive participatory experiences and trust in various actors. This is particularly true of trust in government (rs = .762, p<.001, n = 426), science (rs = .556, p<.001, n = 426) and the media (rs = .741, p<.001, n = 426), with the correlation in each case

[14] On a scale of 1 to 4, where 1 = very positive, 2 = somewhat positive, 3 = somewhat negative and 4 = very negative.

[15] On a scale of 1 to 4, where 1 = very willing, 2 = somewhat willing, 3 = somewhat unwilling and 4 = very unwilling.

corresponding to a large effect following Cohen (1992). In the case of trust in business (rs = .402, p<.001, n = 426) the effect is moderate, per Cohen's scale (1992). Our analysis likewise confirms hypothesis 3.2, showing the correlation between positive participatory experiences and expectations of fairness to be highly significant (rs = .418, p<.001, n = 426). The effect in this instance is moderate, following Cohen (1992). These findings indicate a strong to moderately strong positive correlation between positively assessed participatory experiences and trust in various actors and institutions.

4.3.6 Qualitative data

4.3.6.1 Case selection

In addition to the representative survey discussed above, the findings presented in this paper are based on a regional case study on green hydrogen acceptance carried out in and around Leipzig, with a specific focus on the town of Grimma and its surroundings. This area was selected for two reasons. Firstly, the study called for a region where initial forays into green hydrogen use have already been undertaken. Central Germany's long-standing tradition as chemical industry hub, its well-developed hydrogen infrastructure, and the local presence of research organisations and companies that have been dealing with green hydrogen for some time therefore made this region an ideal candidate, with local industrial initiatives to introduce or expand green hydrogen use in and around Grimma[16,17] providing an additional argument for focussing on this area in particular. Secondly, the research project that forms the basis of this paper was conducted in cooperation with a network of local organisations working towards turning the area into a hydrogen model region[18] – which made selecting interviewees and workshop participants significantly easier.

4.3.6.2 Methodology

Qualitative methods aid in understanding and explaining complex social interactions and relationships (Merriam 2002; Tracy 2010; Upham, Oltra, and Boso 2015). Insofar as qualitative research is primarily geared towards in-depth understanding (Boddy, 2016), the findings from our regional case study are not themselves directly

[16] Cf.: https://www.lvz.de/Region/Grimma/Grimma-soll-ein-Wasserstoff-Standort-werden

[17] Cf.: https://www.mdr.de/wissen/wasserstoffzug-leipzig-grimma-100.html

[18] https://www.hypos-eastgermany.de/en/

generalisable, yet allow for certain generalised conclusions when viewed in combination with the quantitative findings from the representative survey (Mayring 2007). In particular, our qualitative findings facilitate a clearer understanding of the respective roles of trust and participation.

First of all, a total of 24 semi-structured interviews were conducted with members of the general public (n=12) and representatives of government (n=4), science (n=2) and business (n=6), so as to cover the broadest possible range of perspectives (Bernard 2018). Potential interviewees were identified and contacted directly (Robinson, 2014) or – in the case of the local general population – additionally recruited by means of a call for participation distributed via local newspapers, social media, posters, and on- and offline networks. The interviews took place between July and September 2019 and in all but two cases were conducted in person. Each interview lasted 60 minutes and made use of one of two distinct interview guides respectively designed for respondents from civil society and interviewees from the other three fields. All interviews were voluntary and were recorded with consent (Esterberg 2002). The recordings were subsequently transcribed (Oliver, Serovich, and Mason 2005) and analysed on the basis of a deductively developed and inductively supplemented codebook consisting of 6 main categories and 33 subcategories (Tracy 2019), using Mayring's qualitative content analysis (Mayring 2010; Mayring and Fenzl 2019) and with the help of MAXQDA software.

Secondly, two workshops lasting six hours each were conducted in December 2019 in Grimma and in February 2020 in Leipzig, with two separate focus groups totalling 51 participants (Bloor et al. 2001; Wilkinson 1998). Variously representing civil society, local government, business and science, and possessing greatly varying prior knowledge of green hydrogen, the members of each focus group were given the opportunity to engage in equitable dialogue through a process of participatory design (M. L. Heidingsfelder, Bitter, and Ullrich 2019; Sanders 2002; Sanders, Brandt, and Binder 2010). Participants were recruited by means of open calls for participation distributed via local networks, on- and offline media, social media, flyers and posters. Although we aimed to achieve as much balance as possible, both groups contained a majority of older participants (aged ~45 and above) and men. The overall aim was to gain a better understanding of different perspectives by means of a suitable workshop-based approach, while facilitating mutual dialogue and the consolidation of a diverse range of views into a shared, socially accepted vision for local green hydrogen use.

4.3.6.3 Results

4.3.6.3.1 General findings

Overall, in spite of the public's self-ascribed limited knowledge of green hydrogen, we were able to observe a high level of general and local approval in the investigated region. Two factors emerged as particularly decisive for acceptance of green hydrogen, namely its perceived environmental and climate friendliness, and its capacity to help address specific local needs (e.g. within the field of mobility) and current challenges (e.g. with respect to structural change and job creation) by utilising the region's distinct capabilities (as long-standing chemical industry hub). In general, the greatly varying levels of trust placed in representatives of government, the media, science and business, and in the institutions and processes accompanying the introduction of new technologies, could be seen to play a major role.[19]

4.3.6.3.2 Participation, trust, and green hydrogen acceptance, part II

In the absence of prior knowledge on which to base their assessment of green hydrogen, participants in the interviews and focus groups were seen to fall back on perceptions of various aspects of their local context. This consisted, firstly, in the activation and transference of more or less explicitly comparable experiences – for instance with infrastructure projects – to the introduction of green hydrogen. This was particularly evident in participants' reported perception of government and industry processes, and the extent to which they trust the various responsible parties. Here, a complex picture emerges. Whereas positive personal experiences have led to a large degree of trust in municipal-level government actors, the same is decidedly less true for all higher-level government institutions and decision-making processes, which tend to be seen as slow, of questionable integrity and dubiously motivated. Companies are generally viewed with scepticism and considered untrustworthy, with the exception of (very) small local businesses. By contrast, the scientific sector enjoys a large degree of trust, almost without exception.

Secondly, regional self-image could be seen to have a substantial impact on acceptance. Participants were seen to deliberately integrate green hydrogen into their

[19] On the basis of the qualitative findings we derived twelve acceptance factors for green hydrogen, which were then operationalised in a practical guide to the implementation of green hydrogen projects. The guide (in German) can be downloaded here: https://www.cerri.iao.fraunhofer.de/de/projekte/AktuelleProjekte/hypos.html

understanding of their region as a traditional chemical industry hub affected by structural change and fighting for a sustainable future. Drawing various distinct connections between the technology and their own region helped participants come to two favourable conclusions regarding local green hydrogen use. Firstly, a very large majority deemed their region well-equipped to use green hydrogen, thanks to its existing infrastructure, prior history, and knowledgeable individuals and organisations. Secondly, as a sustainable energy technology, green hydrogen was identified as a means of solving important local challenges, with respect to the transition to green energy, job creation and mobility. Thus, in addition to subjective factors, local context, as well as the varying extents to which past experiences have led people to trust different actors, institutions and processes, could be seen to have a major impact on green hydrogen acceptance in the investigated region.

Furthermore, most participants in the case study were seen to consider participation an important and effective tool for promoting social acceptance. Particularly in the run-up to specific planning or implementation projects, it is seen as a means of informing the public and other stakeholders, and of facilitating direct communication between different actors. Participation is thus considered a valuable means of disalienating existing structures and breaking down occasionally hardened fronts between government, business, science and the media on the one hand, and civil society on the other. The inclusion of representatives of the scientific sector is seen as particularly important, given their role as neutral and knowledgeable parties to be consulted in case of uncertainty. Moreover, engaging in direct dialogue with government and civil society is seen as a way for research organisations and businesses to fulfil their respective social responsibilities.

Certain differences between the two focus groups showed participation to be especially conducive to green hydrogen acceptance when people share a common motivation and willingness to become involved (a "reason why"). This not only points to the significance of group composition and dynamics to participatory processes (Scheidler and Pfaff 2019), but also highlights two key factors for the successful promotion of acceptance. Firstly, the methods employed in guiding participatory processes, and the quality of the resulting process itself, are fundamental to overcoming multiple hierarchies and asymmetries (e.g. those resulting from different levels of knowledge) so as to facilitate equitable dialogue. Secondly, inasmuch as acceptance depends on sustained and personally meaningful involvement, it is important to approach participation from a long-term perspective. In addition, transparency and honesty

regarding the potential outcome and limits of any given participatory process are key to avoiding disappointment, which may otherwise adversely impact acceptance.

Taking the issue of trust into account, the focus group workshops were methodologically designed to enable participants to find a common language for addressing complex questions related to local green hydrogen use. Through a carefully guided process employing methods from the field of participatory design, a space was created for participants to get to know a variety of perspectives and arguments, build mutual understanding, and develop a shared vision. In addition, having a neutral research organisation in charge of recruitment and moderation helped make the involved parties more open to the process and to cooperation with one another. Their joint development of a shared vision for regional green hydrogen use could thus foster mutual trust, while enabling especially members of the public to understand and contribute to an otherwise opaque and impenetrable decision-making process.

Participants in both workshops expressed the need for continued dialogue over a longer period of time. This shows that whereas short-term participation can create mutual understanding, trust can only be sustainably established via iterative or longer-term participatory processes. Here, however, a certain paradox arises: while participation can ultimately strengthen trust, the willingness to get involved in participatory processes and engage with other perspectives presupposes a certain measure of trust. Accordingly, despite our inclusive approach and efforts to keep the barrier to entry as low as possible, people who expressed little to no trust in the first place proved difficult to recruit for our study.

Overall, our findings show that given the development and proper implementation of appropriate methods, participation has the capacity to lay the groundwork for trust, which is then to be further cemented over the medium and long term.

4.3.7 Discussion

4.3.7.1 Trust is crucial for green hydrogen acceptance

Our findings reveal trust in actors and institutions responsible for ensuring the fair distribution of costs and benefits to be a significant acceptance factor for green hydrogen. In particular, a relatively strong positive correlation can be observed between acceptance and trust in science. This can be explained by people's tendency to think of green hydrogen primarily as a product of scientific research, when tasked with evaluating it in the absence of a particular practical application and corresponding context. The absence of a specific use case in our study likewise explains why acceptance could be seen to correlate more weakly with trust placed in media outlets to provide impartial information, and with trust in government actors. Given the latter's decisive role in the implementation of new technologies, we expect trust in government to become a far more significant factor in specific use cases. Similarly, the insignificant correlation between local acceptance and trust in the business sector can be explained by the fact that the public does not associate (green) hydrogen with particular companies – and unlike Emmerich et al. (2020), we did not make use of any hypothetical scenarios capable of triggering such associations. The weak correlation between expected distributive justice and acceptance can be explained by the public's lack of experience to date with the costs and benefits of green hydrogen use. Our data nonetheless indicates the existence of a fundamental correlation, which we expect to gather in significance as public experience increases.

As our qualitative findings confirm, only the scientific sector enjoys a large measure of trust. By contrast, little trust is placed in the media, and people tend to be sceptical or very critical of government and business. Our qualitative findings furthermore reveal trust in government and business actors to be a relevant insofar as they are held responsible for properly weighing costs and benefits in terms of sustainability and the common good, and deciding on an implementation strategy befitting the self-image of the region in question. This is in line with Devine-Wright's (2011, 2009) thesis on the significance of place attachment and place identities to acceptance, and highlights the importance of attaining a symbolic fit between place and technology. The latter was particularly noticeable in our case study, where participants could be seen not only to think in terms of positive user experiences and cost savings, but also to explicitly and productively embed green hydrogen in their vision and understanding of community, belonging and regional identity.

We are thus able to qualitatively expand upon our quantitative findings on the role of trust. What emerges as relevant is the extent to which actors and institutions are trusted to manage the introduction green hydrogen in harmony with the local prioritisation of regional self-image and social values such as fairness (cf. Wolsink 2007), as opposed to merely acting in self-interest.

The results of our study build on earlier findings on the role of trust in acceptance in the context of geothermal energy (Pellizzone, Allansdottir, and Manzella 2019; Pellizzone et al. 2015) and wind energy (Aitken 2010; Wolsink 2007; Wolsink and Breukers 2010), and highlight the importance of trust in the relevant institutional framework (Di Ruggero 2014; Scherhaufer et al. 2017; Wolsink 2000, 2007). This encompasses more than just social trust (Siegrist and Cvetkovich 2000) as a means of dealing with personal lack of knowledge (Flynn, Ricci, and Bellaby 2013; Midden and Huijts 2009; Montijn-Dorgelo and Midden 2008); rather, it is above all a matter of trust in the actors and processes that determine how green hydrogen is introduced and used. Our findings furthermore confirm the importance of trust and fairness at community level as described by Wüstenhagen et al. (2007), as well as Mumford and Gray's (2010) findings on the impact of low levels of trust in companies within the energy sector. In addition, our findings echo those of Pellizzone et al. (2017, 2015), which reveal a "deep-seated distrust in institutions, companies and decision-makers" among the general population, while identifying questions of fairness and the common good as decisive for acceptance. Scherhaufer et al. (2017) likewise emphasise the importance of fairness, placing it in connection with policy cores, which are unlikely to change and in turn guided by so-called deep core beliefs – an example of which would be a region's self-image, which can be described as a deep core belief about its values.

4.3.7.2 The quality of participatory processes is important for acceptance

Although our chosen research design does not allow for causal conclusions, a remarkable correlation can nonetheless be observed between green hydrogen acceptance and participation. Individuals with participatory experience display significantly higher levels of acceptance, from which we infer that in addition to openness, environmental consciousness and knowledge, a crucial role is played by people's perceived capacity to influence processes and decisions that affect their personal circumstances (e.g. the introduction of a new technology). Through

participation, people experience self-efficacy and gain a better understanding of different attitudes and needs.

The correlation of favourably assessed participatory experiences with acceptance is particularly interesting. It shows, firstly, that it is not the mere fact of participating that is important, but rather the (perceived) quality of the participatory experience itself – measured not only in terms of methods and results, but also such aspects as the attitude of the organisers and whether participants feel respected. Secondly, positively assessed participatory experiences are specifically conducive to *active* support, which is vital to social acceptance of renewable energy technologies in Germany (Hildebrand, Rau, and Schweizer-Ries 2018; Local Energy Consulting 2020) (cf. 3.2.2).

The findings from the focus groups furthermore show that sound methodology and properly conducted processes are key to enabling successful deliberation and exchange of different perspectives, the development of a common language and – if need be – the reconciliation of different interests. This highlights the necessity of developing and testing sound methodological approaches in order to optimise participatory processes, as pointed out by Scherhaufer et al. (2017) in their discussion of the effectiveness of visualisation. Building on Lienhoop's (2018) argument that different levels of participation are needed along with "substantial improvements so as to enhance procedural justice", our findings underline the importance of the quality of participatory processes, as measured not only in terms of their end results, but also in terms of procedural fairness, and transparency with respect to the potential outcome and impact of the process in question. Overall, our study corroborates earlier findings on the role of participation in acceptance with respect to other forms of renewable energy (Aitken 2010; Breukers and Wolsink 2007; Hildebrand, Rau, and Schweizer-Ries 2018, 2012; Pellizzone et al. 2015) and the significance of fairness in participation (Baxter 2017; Liebe, Bartczak, and Meyerhoff 2017; Wolsink 2007), while expanding on the importance of methodological soundness and the quality of participatory processes and experiences. Building on existing research on the negative impact of alibi participation on acceptance (N. Brennan and Van Rensburg 2016; Langer et al. 2018; Schweizer-Ries et al. 2010), it would at this point be possible to continue a micro-level investigation of particular participatory processes with reference to theoretical (e.g. Schroeter et al. 2016) or practice-oriented (e.g. Australian Public Service 2020; DIALOGIK 2017) efforts to satisfy the demand for increased participation enshrined in official policy.

4.3.7.3 Recurring positive participatory experiences promote trust

The quantitatively calculated positive correlation between participation and trust is likewise confirmed by our qualitative findings. The strong correlation between favourably assessed participatory processes and professed trust in government, science and the media shows that inclusion in planning and decision-making processes enables people to gain an understanding of different perspectives, decisions and compromises, while a positive perception of said processes can help strengthen their trust in the relevant actors and institutions. Conversely, those professing higher levels of trust are more likely to form a favourable assessment of their participatory experiences.

The smaller effect observed in the case of trust in companies can be explained by their tendency to withdraw from the spotlight in predominantly government-steered participatory processes marked by conflict. Combined with the fact that people with high levels of trust in business are likewise less likely to take part in government-steered, public participation processes, this means that companies benefit less from the positive impact of trust.

The significant correlation between positively assessed participatory experiences and expectations that the costs and benefits of green hydrogen usage will be distributed fairly similarly shows that the methodological design of a given participatory process affects participants' perceptions and expectations of fairness. Conversely, people with high levels of trust in institutions responsible for upholding distributive justice are more likely to describe participatory processes as successful.

Our qualitative findings also provide further evidence of the paradox mentioned in section 4.3.2 above. For the most part, the people we managed to recruit for our interviews and focus groups displayed at least some measure of basic social trust. Individuals whose trust had been largely or fundamentally eroded, however, were far less receptive to our recruitment strategies and proved difficult or impossible to include in the process.[20] Nevertheless, by providing the necessary framework and methods for equitable dialogue and mutual understanding, the participatory process could be seen to have a positive impact on participants' trust. Although this shows that participation can promote trust, we maintain that a certain measure or minimum level of trust should not become a prerequisite for inclusion in participatory processes, lest they become incapable of reaching those members of the public who are more given

[20] For instance, out of nearly 50 people who were directly contacted via e-mail and/or telephone, only 12 ended up taking part in the process. The remaining participants were recruited via an open call for participation distributed via posters, flyers and mailing lists.

to distrust. This would in turn not only serve to widen the social gap between the more sceptical and more trusting segments of society, but may also make widespread social acceptance impossible to achieve.

The bilateral, interwoven relationship between trust and positive participatory experiences makes their mutual reinforcement possible, especially over a longer period of time. Accordingly, as recurrent positive experiences can capitalise on and strengthen existing trust, the design of participatory processes needs to adopt a long-term perspective. This is confirmed by the desire expressed in our study – by both interviewees and focus group participants – for continued dialogue and for joint deliberation to take place on a more permanent basis. Participation should thus become systemic and be organised on a recurring or long-term basis, so as to effectively build trust and avoid the risk of simply preaching to the choir.

Against this background, we argue that social acceptance should be understood as a matter of responsible innovation (Stilgoe, Owen, and Macnaghten 2013), in acknowledgement of the need to open the innovation system to civil society. This can be achieved by means of long-term public participation. By building trust over the long term and promoting acceptance at an early stage, responsible innovation's central tenet of democratisation can be embedded in the development and use of green hydrogen technology.

First of all, we maintain that approaching acceptance as a matter of responsible innovation entails a stronger focus on systemic measures, so as to firmly cement the role of social values in steering green hydrogen use by means of cross-sectoral participation – and in particular through the involvement of civil society. This means taking the general public's knowledge seriously and affording the public as actor more trust, responsibility and authority – which in turn calls for research into diverse forms of interaction and dialogue.

Secondly, there needs to be a stronger focus on innovation processes. Efforts and strategies to create conditions amenable to acceptance should look not only towards the subject, but also the object of acceptance and its surrounding context. It does not suffice to try to determine which individual factors – for instance psychological characteristics – tend to favour acceptance. Instead, we argue that technology acceptance also requires investigating and (re-)shaping research, development, planning and implementation processes, with the resulting participatory innovation processes providing a space for different forms of practical involvement and the expression of a variety of reactions. Acceptance is thus not treated as a given outcome;

instead, as particular practical solutions based on green hydrogen technology are collaboratively developed and implemented in response to specific needs, the relevant acceptance criteria are at the same time collaboratively inscribed in the solutions themselves.

Adopting a participation-oriented understanding of acceptance as a matter of responsible innovation could serve to address numerous limitations of acceptance research – which has been criticised for not affording the public enough trust (Aitken 2010), not paying enough attention to different forms of (expressing) approval or rejection (Batel 2018; Batel, Devine-Wright, and Tangeland 2013), using predominantly quantitative approaches that potentially rely on a simplistic understanding of acceptance, thus overlooking important aspects thereof (Batel 2018; Flynn, Ricci, and Bellaby 2013), displaying a pro-acceptance bias in treating acceptance as both given and normatively desirable (Hagen et al. 2018), and generally inadequately factoring in the relevant institutional and sociopolitical context (Di Ruggero 2014; Scherhaufer et al. 2017; Wolsink 2000). By abandoning a top-down perspective (Batel, Devine-Wright, and Tangeland 2013) in favour of a participatory governance framework predicated on the systemic integration of civil society, more trust is placed in societal actors and in their knowledge, the focus is shifted to the institutional context surrounding innovation, and a greater variety of reactions is made possible via early participation – all of which serves to broaden the formerly narrow concept of social acceptance. Thus our findings reveal the limits of a passive understanding of acceptance that treats the acceptance object as given (Upham, Oltra, and Boso 2015), while highlighting the sociopolitical dimension of innovation and acceptance (Braun and Griessler 2018).

4.3.8 Conclusion and policy implications

4.3.8.1 General implications

In order to investigate green hydrogen acceptance in Germany, we chose a mixed-method approach, combining quantitative data from a representative survey with qualitative data derived from interviews and focus group workshops. Given its key role in sector coupling and the establishment of a sustainable energy system, green hydrogen is central to the process of social transformation accompanying Germany's energy transition. Since practical implementation of this technology is still at an early stage, early social acceptance can be promoted via responsible innovation and rollout processes, thereby avoiding the kinds of conflicts and delays seen in the case of other renewable energy technologies.

Overall, our findings show the public to be simultaneously very unfamiliar with and very open towards – and in some cases even extremely interested in – green hydrogen, primarily thanks to the associations and promise of sustainability it carries. Trust is seen to be essential for acceptance: not only trust in science, government and the media, but also in institutions responsible for upholding regional values and ensuring the fair distribution of costs and benefits. What emerges as decisive in this regard is the extent to which the public is confident that green hydrogen will be used in accordance with regional values, demands and capabilities.

Secondly, our findings show participation to be an effective instrument for promoting acceptance in general and active support in particular, and that positive participatory experiences can play an important role in fostering trust. Creating such experiences requires the careful, scientific development and implementation of suitable participation formats and methods, which take the relevant sociopolitical context and participants' previous experiences into account, and which facilitate open and long-term dialogue. Simply ensuring that participation takes place is not enough: when the goal is to build trust and promote acceptance, non-positive or negative participatory experiences can exact a very high price (N. Brennan and Van Rensburg 2016; Langer et al. 2018). Accordingly, we conclude that instead of being treated as a mere tool or short-term measure, participation should be instituted over the long term with a view to building and strengthening trust.

Against this background, we argue that social acceptance of green hydrogen should be approached from the perspective of responsible innovation, and participation understood as a means of democratising innovation by systemically altering the relationship between civil society on the one hand, and government, science and

business on the other. In this way, trust can be strengthened on a structural level, thus staving off potential conflicts in the medium and long term.

4.3.8.2 Policy implications

Our study has revealed instances of significantly eroded public trust. In addition, the window of opportunity for cementing public positivity and openness towards green hydrogen is closing with increased use of the technology. Furthermore, any conflicts that happen to draw the public's attention risk creating a public image that can be difficult to change after the fact. All of these factors contribute to a certain level of urgency.[21]

Our findings carry certain implications both for innovation policy in general, and with respect to green hydrogen in Germany. In terms of the latter, we offer two sets of practical recommendations. Firstly, there is a need for measures that increase public awareness and ensure widespread basic knowledge of green hydrogen (Bögel et al. 2018). It is important that the information be presented in an accessible format, so as not to exclude especially older segments of the population and those with lower or no academic qualifications. Special attention should be paid to transparency and a balanced presentation of the technology's advantages and disadvantages, current state of development and future possibilities, so as to avoid creating false expectations regarding sustainability. Secondly, advocates of green hydrogen need to be activated by means of participatory measures that target the emotional dimension of environmental awareness, for instance via emotionally engaging visions for regional hydrogen use. This requires creating positive participatory experiences, and evaluating their impact in a scientifically rigorous manner.

Our findings furthermore support three broad recommendations for innovation policy. Firstly, there is a need for substantial investment in systemic trust in government, science, the media, and institutions that safeguard distributive justice, so as to create a stronger basis for acceptance of new technologies and the transition to green energy as a whole – all of which is to be understood as falling under the remit of innovation and technology policy. Secondly, we recommend more clearly acknowledging the sociopolitical nature of innovation by understanding questions of

[21] The federal government seems to have at least recognised the urgency and critical importance of social acceptance. The most recent amendment to the Renewable Energy Sources Act, which plays a central role in Germany's energy transition, lists "acceptance of the further expansion of renewable energies" as one of the six most important current issues (Deutscher Bundestag 2020).

trust and fairness as integral parts of innovation policy, which should systematically inform the shape of innovation processes – for instance via the establishment of particular funding criteria or standardisation processes. Lastly, it is necessary to keep exploring and testing ways of further structurally democratising the innovation system via participation – not only within academic research, but also for instance as part of municipal innovation projects.

4.3.8.3 Limitations and avenues for further research

Despite combining a representative survey with qualitative data, our study faced certain limitations. First of all, as our investigation of social acceptance of green hydrogen took place at a very early stage of its deployment, we did not have the option of examining any existing implementation projects – and could thus only focus on the technology's potential use in participants' local surroundings. Secondly, our particular focus on a region with a long history in the chemical industry resulted in qualitative findings that are only generalisable in combination with the quantitative results of our representative survey. Nonetheless, as the combination of quantitative and qualitative methods proved useful in exploring the complexities of social acceptance, we encourage further studies using a similar mixed-method approach.

Our findings indicate, firstly, that the relationship between institutional trust and regional values merits further investigation. This could aid in translating broad regional energy concepts into specific strategies or business models (Kalkbrenner, Yonezawa, and Roosen 2017). Secondly, suitable criteria for assessing participatory processes in terms of their impact on acceptance need to be developed and appropriately evaluated. Thirdly, it is worth investigating the impact of various forms of recurring and long-term participation on institutional trust over a longer period of time. Fourthly, further research needs to specify participation in terms of its timing in research, development and innovation projects, the responsibilities attributed to different stakeholders, and the knowledge required to initiate such processes. Finally, our attempt to conceptually and empirically interlink acceptance research and responsible innovation gives rise to the following important questions. How can responsible innovation benefit from the scientific and methodological expertise of acceptance research? Which forms of systemic participation are particularly conducive to acceptance? What roles and processes would exist in a more democratic innovation system geared towards trust and acceptance, and which systemic adjustments would the establishment of such a system require?

Funding

This research was funded by the German Federal Ministry for Education and Research (Bundesministerium für Bildung und Forschung (BMBF)) under grant number 03ZZ0746.

Acknowledgements

The research project that forms the basis for this paper would not have been possible without significant contributions by K.G. and M.K. For this we would like to express our heartfelt gratitude.

5. Discussion: On the prospects of responsible innovation

This chapter first summarises the main findings of the three articles included in this dissertation before discussing them with regard to the future of responsible innovation in general and the role of participation in particular. The presentation of the most important results focuses on the findings with regard to the research questions formulated in 3.2. The more in-depth presentation of the individual results as well as their detailed discussion can be found in the respective articles (4.1, 4.2, 4.3). Finally, the research contribution made by this dissertation is highlighted and important limitations are pointed out. Overall, the aim of this chapter is to summarise the results of the three articles and, as far as possible and appropriate, to bring them together into an integrated overall picture of the status quo of responsible innovation against the background of current research.

5.1 Results

This chapter summarizes the most important findings of the three case studies. As this section aims to bring together the findings of three individual publications, it cannot include all of their specific findings. Rather, the following subsections provide a rough overview of the argument (theoretical case study), the developed approach (conceptual case study), and of the for this context most important results (empirical case study).

5.1.1 Responsible innovation and normative economics

The first article investigates the implications of the normative complexity of innovation from a perspective of economic theory for the concept of responsible innovation (Häußermann and Schroth 2020). To this end, it first introduces the concept of responsible innovation as a means of aligning innovation with ethics. In the brief introduction to the emergence and main characteristics of responsible innovation, the importance of participation and the role of individual citizens is particularly highlighted. Through the participatory ambition to involve societal actors and individual citizens in the development and introduction of innovations as early as possible, "responsible innovation contributes to a new distribution of moral labour and, by extension, (co-)responsibility for innovations" (Häußermann and Schroth 2020, 351). While the societal evaluation of innovation is usually seen as organised via a market test (cf. chapter 2.2.), the approach of responsible innovation aims to shift ethical responsibility more towards the forefront of the innovation process, i.e. to focus on the role of the producers of innovations. The article argues that, first, the conceptualisation of the ethical evaluation of innovation from the perspective of mainstream or neoclassical normative economics is intrinsically problematic and incomplete, and responsible innovation therefore too hastily neglects a substantiating reasoning from the economic perspective. What the article shows instead is that an account of normative economics based on the criterion of preference learning is more plausible, especially for the context of innovation. This ultimately allows economic perspectives to be taken into account more thoroughly by responsible innovation and to be included in a conceptualisation of the concept. Because insofar as the normative criterion of preference learning can avoid the problems of the classical account of mainstream economics based on the satisfaction of individual preferences, a procedural conceptualisation of welfare as a new opportunity to discover individual preferences enables both a theoretical underpinning from the perspective of economic theory and a functional expansion of practical approaches to implementation. In short,

“the account of preference learning that we are propounding focusses on how individuals actually develop their preferences or, in other words, how they ‘learn’” (Häußermann and Schroth 2020, 356). As the normative criterion for innovations from an economic viewpoint can no longer be merely individual purchasing choices or a so-called "market test", but rather the individual learning of preferences, attitudes, desires, concerns, or rejections in the face of innovations (new technologies, business models, products, services, etc.), a new dimension of the redistribution of responsibility and of a new division of moral labour through responsible innovation emerges. The rejection of the individual level in the sense of an evaluation by individuals as consumers must therefore not result in a neglect of individual evaluation as such, but rather be replaced by a focus on the possibilities of individual learning. In light of this, responsible innovation also means to give individuals the means to learn: “that is, to form (ethical) preferences towards innovations and reflect upon their evolving preferences with respect to, for instance, emerging technologies” (Häußermann and Schroth 2020, 357). To this end, as the article shows, an approach of "laboratories in real-world contexts" or also called living labs, which is already receiving attention in the context of responsible innovation for other reasons, is a very suitable approach. While typically a better understanding of the complex environment of new technologies and the interactions between users and individual applications stand in the focus, this practical approach can also be used to give individuals the opportunity to try out innovations, to experience opportunities, risks, and potential impacts and thus to develop their own attitudes and preferences.

In a nutshell, the most important results of the first article for the present discussion can be summarised as follows:

- The normative perspective of standard welfare economics based on the satisfaction of preferences is problematic and insufficient for the context of innovation.
- Instead, an alternative normative criterion based on preference learning can be derived from the discussion of economic theory.
- A normative approach based on preference learning allows responsible innovation to take into account perspectives and current findings in the field of (normative) economic theory.
- Based on the concept of preference learning, an approach to responsible innovation inspired by economic theory both provides a novel theoretical substantiation as well as an expansion of its practical implementation.

- Rather than conceptualizing individuals as mere consumers of innovations, responsible innovation thus requires enabling individuals to form preferences and evaluate innovations, so as to align innovation with ethical demands and so that they can fulfil their (normative) role as co-innovators.

In response to the research question presented in section 3.2.1. above, the following answers can be given. First, a normative approach based on standard welfare economics encounters at least two problems in the context, an epistemic and a behavioural problem, which is why a normative conceptualisation based on the criterion of preference learning seems more appropriate. For responsible innovation, it follows that integrating a normative economics perspective can be valuable both on a theoretical and practical level, as the role of individual evaluation can be salvaged less than individual consumers' market-based choices but as preference learning. As a result, responsible innovation requires to enable and foster individuals' learning and reflection capacities through participation, e.g., by adjusting and suitable conditions and institutions for producing, marketing, and using innovations.

5.1.2 Responsible innovation and business ethics

The second published article examines the intersection of responsible innovation and business ethics using the example of AI technologies and the debate on AI ethics that has emerged in recent years (Häußermann and Lütge 2021). To this end, I first summarise the backdrop and the current state of the AI ethics debate and identify five current shortcomings. According to this, the current debate is flawed by the fact that it (1) neglects the business context of developing and employing AI systems in society, (2) is biased toward a technological solutionism, (3) succumbs to an individualist focus, (4) is problematic in its implementation and lacks accountability and clear impact, and (5) lacks a clear relationship to legal regulation. In light of these shortcomings, the article shows how responsible innovation can integrate established theories and concepts of business ethics through the unique role of participation. With this aim, building on the outlined status quo of AI ethics, the business ethics theory of order ethics is introduced and applied to the context of AI for the first time. This reveals that order ethics as a contractualist theory of business ethics offers a suitable business ethics supplement for the field of AI and innovation as a whole, since on the one hand it focuses on the examination of different values and value conflicts in pluralistic societies, and on the other hand it takes into account the entrepreneurial practice of markets and given competitive conditions. Accordingly, an extension of responsible innovation based on order ethics to include a business ethics perspective seems to be the appropriate basis for operationalising a participatory governance framework. Building on the importance of rules as a justified level for addressing ethical conflicts, the article then develops a deliberative governance framework that requires the participation of all affected stakeholders in order to obtain rules that are acceptable to all. Based on the normative criterion of mutual advantages for all actors affected by the rules, a procedural extension of order ethics is developed which uses the example of AI to show what role participation can play for responsible innovation, particularly against the background of established theories of business ethics. Finally, this not only broadens the perspective on participation, but also allows the framework of responsible innovation to be extended to other phases of value creation. By building on mutual advantages for the stakeholders concerned, the deliberative governance framework based on order ethics ultimately aims to enable business practices in the sense of pluralistic value creation. In this way, a business ethics complement to responsible innovation helps to take into account the commercialisation and marketing of innovations by companies using a participatory approach. The article therefore not only succeeds in showing a way of conceptually integrating responsible

innovation with established theories and concepts of business ethics, but also in making the potential of responsible innovation useful for later phases of the innovation process, i.e. the market-based distribution of innovations, in the sense of pluralistic value creation.

In a nutshell, the most important results of the second article for the discussion here can be summarised as follows:

- Beyond the example of AI technologies, order ethics offers a suitable basis for combining a business ethics perspective with the approach of responsible innovation.
- The concept of a participatory governance framework provides the appropriate conceptual interface for integrating responsible innovation with business ethics considerations.
- An extension of responsible innovation to include a business ethics perspective based on a procedural account of order ethics allows aspects of a market-based distribution, which is central to innovation, to be taken into account in terms of pluralistic value creation.

Against this background, the research questions formulated in section 3.2.2 may be answered. The contractualist theory of order ethics offers a suitable framework for developing a participatory governance framework for innovation, insofar as the relevance of rules and the normative criterion of mutual advantages provide the basis for the inclusion of all stakeholders concerned. Secondly, it has been shown that normative-ethical aspects of a new technology such as AI can also be conceptualised to a significant extent as entrepreneurial challenges, or that this is even necessary in order not to lose sight of essential aspects (Häußermann and Lütge 2021, 4–5). The article shows how a participatory governance framework based on a procedural extension of order ethics is suitable for responsible innovation to take this normative dimension of innovation into account. As a result, responsible innovation can better reflect the organisational perspective of firms and the commercial context innovation and be better aligned with the objective of pluralistic value creation.

5.1.3 Responsible innovation and social acceptance

The third article looks at the relationship between responsible innovation and social acceptance in the context of an empirical case study (Häußermann et al. 2021). Based on the investigation of the social acceptance of green hydrogen in Germany using both qualitative and quantitative data, the article focuses on the role of participation and resulting implications for the relationship between responsible innovation and the social acceptance.

The quantitative results show a positive correlation between participation and the acceptance of green hydrogen, especially with regard to the willingness to actively support the adoption of green hydrogen technologies. Our statistical analysis reveals significantly higher levels of approval of green hydrogen among people with participatory experience (mean 1,69, mean rank 874,73) compared to those without (mean 1,94, mean rank 1089,70). Moreover, people who assess their participatory experiences favourably display significantly greater willingness for active involvement (mean 1,88, mean rank 147,49) than those who do not (mean 2,31, mean rank 192,18). These findings indicate a notable correlation between participation and approval of green hydrogen, and between the quality of people's participatory experiences and their willingness to become actively involved in promoting its use. Furthermore, the results point to a significant correlation between participation and trust in government, business, academia, and the media. Our results reveal a highly significant correlation between positive participatory experiences and trust in various actors. This is particularly true of trust in government ($r_s = .762$, $p<.001$, $n = 426$), science ($r_s = .556$, $p<.001$, $n = 426$), and the media ($r_s = .741$, $p<.001$, $n = 426$), with the correlation in each case corresponding to a large effect following Cohen (1992). In the case of trust in business ($r_s = .402$, $p<.001$, $n = 426$) the effect is moderate according to Cohen's scale (1992). These findings indicate a strong to moderately strong positive correlation between positively assessed participatory experiences and trust in various actors and institutions.

A similar picture emerges from the qualitative results on the relationship between acceptance and participation. Here, not only a nuanced and complex relationship of trust between citizens on the one hand and different institutions on the other was revealed, but also the impact on the acceptance of a new technology such as green hydrogen was shown. Participation was revaealed to be an effective instrument for promoting both the social acceptance of the technology and trust in the actors involved. All in all, the qualitative and quantitative results show firstly the great importance participation can have for promoting the acceptance of new technologies.

Secondly, they draw attention to the role that trust in actors, institutions, and processes plays in acceptance and that it can likewise be promoted through participation. In particular, the article shows that the quality and design of participation processes have an influence on acceptance and that recurring or long-term participation is suitable for promoting trust.

This entails a number of implications for responsible innovation. It shows that the elementary focus on participation can be an effective instrument for promoting acceptance of new technologies and innovations and can also contribute to increasing trust which is vital for innovation. Thus, understanding social acceptance as a matter of responsible innovation implies two key insights for the latter concept. First, responsible innovation should make an important contribution to the systemic strengthening of trust in the innovation system and its actors, institutions, and processes. In this sense, responsible innovation can not only achieve an alignment of societal needs and values on the level of particular technologies, but also pursue more integrated strategies and systemic goals. Sustainable results at this level, such as strengthened systemic trust, can in turn have a positive impact on particular projects. Secondly, responsible innovation may consider a stronger focus on the design of innovation processes and the role of participation. Responsible innovation not only systematically embeds participation in the innovation process, but also enables a wide spectrum of opportunities for participation and involvement, and can thus also reflect a wide range of reactions. In this way, numerous limitations of acceptance research can also be addressed.

In a nutshell, the most important results of the third article for the discussion here can be summarised as follows:

- Participation can be an effective tool for promoting the acceptance of new technologies and strengthening the trust in actors, institutions, and processes that is relevant for innovation.
- Social acceptance should be understood as a matter of responsible innovation which entails several advantages.
- Conceptualising social acceptance through the lens of responsible innovation opens up the space for very different reactions and relationships between innovations and citizens – and thus allows to go beyond a passive understanding of acceptance.

- The importance of trust also underlines the systemic relevance of responsible innovation not only to improve particular solutions, but also to contribute to a democratisation of the overall innovation system.

With this, the research questions formulated in section 3.2.3 can be answered successfully, insofar as (1) the positive impact of participation on acceptance as well as (2) the implications and advantages of a conceptualisation of acceptance as a matter of responsible innovation can be shown. As a result, the article can not only shed light on the relationship between social acceptance of new technologies and responsible innovation, but also points out avenues for a conceptual enhancement and extension of responsible innovation.

5.2 Discussion

In this chapter, the results of the three individual articles that have been briefly summarised above will be discussed with regard to the concept of responsible innovation and contextualised against the background of current research. The aim is to present an integrated conceptual conclusion in view of the future of responsible innovation. To do this, the individual results of the articles as presented and discussed in greater detail in the respective sections of chapter 4 will not only be substantially summarised, but also generalised, and sharpened in order to derive a clear conclusion of this dissertation.

A conceptualization of responsible innovation based on economic theory leads to a revived interest in individuals and their capacities to ‚learn'

The results of the first article not only show that an examination of the normative complications of innovation is valuable in terms of economic theory, but that this can also offer added value for the concept of responsible innovation. However, except for some single exceptions (cf. Schlaile et al. 2018; Hühn 2018), the notion of responsible innovation has not yet been addressed from the perspective of normative economic theory. A recourse to the normative criterion of preference learning can provide responsible innovation with both theoretical underpinnings and important practical starting points. Since markets do not yet exist for many innovations and individuals have not yet developed preferences, a normative orientation towards the latter seems at least problematic (Binder 2013). But instead of completely rejecting the view of individual evaluations in the sense of economic theory and instead of the satisfaction of preferences through market-based choices, a procedural view of the learning of preferences should come into focus (Dold and Schubert 2018). For responsible innovation, this means that participation should be understood in a more bidirectional way and that the aim of inclusion should not be confined to the improvement and co-creation of innovations but include the facilitation of learning in the sense of the development of preferences. In other words, part of responsible innovation should also be helping individuals to understand their own needs and desires and to compare them with the opportunities, risks, costs, and benefits as well as possible side effects. It is plausible that corresponding support in the individual formation of preferences will take place along all phases of the innovation process and by means of very different formats and methods: From the more integrated development of visions and desires for the future, to targeted pedagogically-inspired interventions for empowerment and

new approaches to product information, to new responsible innovation approaches in marketing and consumer protection. Thus, new fields of action and tasks for responsible innovation are coming into focus, which aim to treat individual preferences not in the sense of outcomes, but to assign a value of their own to the diverse and complex processes of learning. In this sense, the task of responsible innovation can be to modify „those contextual elements that foster individuals' means of preference learning, viz., educational and cultural institutions that provoke individuals' reflection of their evolving preferences and access to markets that allow individuals to test different preferences on a continuous basis" (Dold and Schubert 2018, 237). In this way, responsible innovation can not only contribute to the responsible design of a "market for preferences" (Earl and Potts 2004) and also address more strongly the behavioural dimensions of innovations (Akerlof and Shiller 2015). The consideration of learning as part of responsible innovation can also be understood from a macro perspective of the societal and economic level. Even though standard economic theories and models are based on growth and consumption, it is only recently that a growing interest in the demand side of markets and the relationship with innovation (policy) can be recognised (Witt 2001; Boon and Edler 2018; Edler 2006; J. Schot and Steinmueller 2018). Insofar as the creation of individual preferences is decisive for the formation of the demand side of markets on the micro level, responsible innovation can start here and help in the formation of responsible preferences and demand structures that take into account major societal challenges. Insofar as control via markets and demand is an important mechanism and driver for social development, responsible preference learning can be an effective instrument for the transformation of markets and societies in the sense of sustainability.

A business ethics inspired approach to responsible innovation leads to a more comprehensive integration of the organisational level of businesses and their role in society

The results of the second article clearly show the potential and added value of connecting responsible innovation with business ethics theories and concepts. Not only can societal and ethically relevant aspects of later phases of innovation processes be given greater consideration, but in particular the role of companies as central actors for the development, diffusion, and use of innovations can be represented better in conceptual terms. Building on the growing attention paid to responsible innovation in an industrial contexts (Dreyer et al. 2017; van de Poel et al. 2017; Martinuzzi et al.

2018) (see section 3.2.2) as well as initial connections of responsible innovation with more traditional theories of business ethics (Brand and Blok 2019), this dissertation focuses on the role of participation from a business ethics perspective. Insofar as responsible innovation understands participation primarily as a governance model for innovation in terms of specific technologies or applications, a business ethics perspective extends the role of a participatory governance framework to the organisational level of companies, their societal role, and how they interact with different societal actors. In addition to a critical reflection on deliberation as a central governance mechanism in an entrepreneurial and market economy context (Brand and Blok 2019), the reference to contractualist theories of business ethics (Luetge, Armbrüster, and Müller 2016; Luetge 2016) allows responsible innovation to be positioned as a complementary approach between the governance level of individual innovations on the one hand and the governance level of laws on the other. By defining rules as a field for ethical action, responsible innovation can offer a strategic instrument to enable a responsible approach to innovation through the participatory inclusion of different actors and thus the opening of companies along the lines of a participatory governance framework. In the sense that responsible business action in the context of innovation does not only include the development of individual technologies or the consideration of legal requirements, responsible innovation needs to include the meso level of business behaviour within society. The proposal for a conceptual extension of responsible innovation presented in the second article contributes to these considerations. In this way, responsible innovation can not only pick up on different discussions in the field of stakeholder engagement and stakeholder dialogue (Brand, Blok, and Verweij 2020; Vincent Blok 2018a) or cross-sector collaboration in the context of new (open) innovation constellations (Carayannis and Campbell 2012b; Chesbrough, Vanhaverbeke, and West 2006), but also draw on other strands of research in the field of business and corporate ethics, such as political CSR (Scherer and Palazzo 2007, 2011). In a similar way as Brand et al. (2020) draw attention to the role of conflicts in terms of an approach of "agonistic deliberation" and Schormair and Gilbert (2020) discuss the handling of conflicts in terms of discursive justification, the approach presented in this dissertation does not aim to contain or bring together different perspectives to a common denominator and a single shared value. Rather, the extension of responsible innovation in terms of the participatory-deliberative governance framework aims at pluralistic value creation that conceptualises business activities as the realisation of different societal values. In this way, the expansion of responsible innovation to include a business ethics perspective aims to integrate normatively relevant dimensions of innovation that are central to a societally

responsible approach. For it is not only the design of technologies and products along societal values, but also how companies then use them, exploit them in business models, and make them part of corporate strategies and visions. Without taking these levels of innovation into account, responsible innovation risks remaining a toothless tiger. Therefore, a comprehensive approach to responsible innovation must also integrate economic and business ethics perspectives in order to be able to shape business practices in the context of innovation in a societally responsible way. The aim of responsible innovation should thus be to enable pluralistic value creation: Innovation as key to value creation must really aim at and realise societal values – even if we first have to be clear about which values these are (Mazzucato 2018).

Understanding social acceptance as a matter of responsible innovation takes the key role of trust into account and paves the way for a democratisation of innovation systems

The results of the third article of this dissertation shed light on the nexus between social acceptance and responsible innovation using the example of an empirical study of green hydrogen in Germany. In particular, the effectiveness of participation in promoting acceptance and the special role of trust were demonstrated. The dissertation can thus make an important contribution to the intersection of technology acceptance research and responsible innovation, which has only been researched to a limited extent to date (Correljé et al. 2015; Koirala, van Oost, and van der Windt 2018; Pellizzone et al. 2015, 2017). On the one hand, this concerns the advantages of viewing social acceptance from the perspective of responsible innovation for current acceptance research and its various shortcomings. For example, the lack of trust in societal actors and their knowledge (Aitken 2010), a limiting view of the variety of different forms of (expressing) approval or rejection (Batel 2018; Batel, Devine-Wright, and Tangeland 2013), or the failure to consider relevant factors of the institutional or socio-political context (Scherhaufer et al. 2017; Wolsink 2000; Di Ruggero 2014) have been criticised. This manifests itself, among other things, in a simplistic understanding of acceptance (Flynn, Ricci, and Bellaby 2013; Batel 2018) and an implicit pro-acceptance bias (Hagen et al. 2018), according to which the acceptance of innovation as approval or support is already presupposed and expected from the outset. In this context, an approach based on responsible innovation offers the opportunity to broaden the view and pursue a less "top-down" perspective on acceptance. This means that by including societal perspectives in the innovation process at an early stage, not only are different reactions and relationships between,

say, a technology and citizens facilitated, but these also feed into the innovation as design factors. On the other hand, this intersection does not only give responsible innovation access to another established field of research (see section 3.2.3), but above all to a central dimension for the practical relationship between innovation and society. Although acceptance has therefore always been a declared objective of responsible innovation (Ribeiro, Smith, and Millar 2017), an expansion or strategic opening of responsible innovation for the topics and methods of social acceptance and technology acceptance research seems necessary. One reason why the concept of responsible innovation has rarely been taken into account in the context of social acceptance research could be that the latter is an established and consolidated field of research – with known limitations. Understanding social acceptance as a matter of responsible innovation, however, allows moving closer to the overarching vision of a more democratic governance of innovation systems (Braun and Griessler 2018) by challenging the at least implicitly widespread passive understanding of diffusion and adoption as well as acceptance of innovations and enabling a more complex understanding of innovation and its use in society. This is characterised not least by an explicitly more active, design- and participation-oriented role of society in the innovation system. It also raises an important question for responsible innovation: How can this understanding be scaled in the sense that it can claim validity for all (potential) users, as in the case of adoption and acceptance – and not only for those who could possibly be actively involved? For strengthening responsible innovation, a greater consideration of acceptance and acceptance research thus means two things: on the one hand, it opens up access to a widespread understanding of the relationship between innovation and society, both in academia as well as in business and policy. Secondly, it is precisely through a shift in this (implicit) understanding of the relationship between innovation and society that responsible innovation can effectively contribute to a democratisation of the innovation system.

Responsible innovation needs to reconquer innovation as a political concept

Against the background of the broader introduction (chapter 2.1.) and the context of the ethical dimensions of innovation in general (chapter 2.2), the results of the three articles of this dissertation point to the overarching challenge of reclaiming innovation as a political concept. Insofar as a technical-economic understanding of innovation as an end in itself and as key instrument for economic growth (see chapter 2), measured as an increase in gross domestic product, has established itself in the course of the 20th

and 21st centuries and continues to be dominant today, it is necessary to reclaim the originally inherently political character of innovation (Benoit Godin 2015; Benoît Godin 2008) in view of the increasing and more profound ethical-normative complications of innovations. This is not primarily meant to indicate a change in the substantive matter, according to which it would be a matter of an explicit thematic orientation or steering of technological change, but rather to reflect the fact that the normative-ethical dimensions and societal aspects are not so much side effects or undesirable consequences but *fundamental elements of the concept of innovation* per se. In this sense, a re-politicisation of innovation is about understanding the ethical dimensions (see chapter 2.2.2) in terms of a comprehensive societal embedding of innovation and to pay as much attention to them as to other technical-economic dimensions. The three articles illustrate this need from different perspectives: The first shows that the standard economic normative evaluation of innovations in the sense of satisfaction of preferences is problematic and should be replaced by a criterion based on preference learning based on the idea of responsible innovation. This means that instead of a reduced economic-technical understanding of the normative evaluation of innovation through purchase decisions, a more comprehensive approach taking into account the manifold processes of preference learning and their socio-political embedding is needed. The second article points to the societal role of companies as innovators. Especially because companies develop and use new technologies and innovations, they need to understand themselves as political actors who actively shape society not merely, say, through the sale of products to citizens as customers. And finally, the third article shows that the social acceptance of innovations should not be understood through a passive top-down conceptualisation but can be effectively achieved through different forms of participation in the sense of responsible innovation. Social acceptance of new technologies is thus not just a question of diffusion and adoption but rather an expression of a complex societal negotiation process about the costs and benefits, opportunities and risks of an innovation. Technology acceptance is thereby an important mechanism for the societal and democratic shaping of technological change. In short, the different normative-ethical dimensions of innovation reveal that innovation is political in the sense that it is an important mechanism for how society envisions itself and evolves. Innovation does not consist of financial growth, nor does it end up in individual purchasing decisions: Innovation is one of the central fields in which societal progress is shaped – a crucially political process. Innovation ethics as a reflection on the normative-ethical dimensions of innovation (see section 2.2.2) is therefore always political and even a political endeavour.

5.3 Contribution to research

In the following, I will examine the overall contribution of the dissertation to current research going beyond the specific contribution of each individual article to its particular area of scholarly inquiry.

First, this dissertation contributes to current research in the field of responsible innovation and in particular to its conceptual development. As a fundamentally inter- and transdisciplinary field, research on responsible innovation brings together a variety of theoretical, empirical, and practice-oriented perspectives (see also section 3.1). However, the starting point of this dissertation was the observation that at three interfaces to neighbouring research strands, the conceptual context and the opportunities for responsible innovation have remained rather unexplored so far. In this sense, an important contribution to research consists in (1) illuminating for the first time the relationship between responsible innovation and normative economic theory, (2) linking responsible innovation with an established theory of business ethics via the distinctive role of the participatory governance framework and for the first time with the contractualist idea of order ethics, and (3) showing the possibilities of a stronger conceptual integration of responsible innovation with acceptance research.

Furthermore, the overview of the different normative dimensions of innovation in terms of a circle of innovation ethics constitutes one of the first attempts at a systematisation of the field (see section 2.2.2). To the knowledge of the author, there is currently no comparable overview that does not focus solely on normative-ethical aspects in the sense of negative side-effects (Biggi and Giuliani 2021) but combines these with other relevant dimensions. Here, the dissertation contributes to embedding the approach of responsible innovation in an emerging field of innovation ethics and to collect and contextualise the various studies and contributions that exist to date on the different normative dimensions of innovation.

5.4 Further research

To the extent that this dissertation connects different research strands and thus explores new territories, a multitude of follow-up questions and starting points for future research emerge. These include, for example, the following questions:

Responsible innovation and (normative) economics:

How can an approach to responsible innovation, i.e., the alignment of innovation with societal values, be anchored in economic theories and models? How could an economic theory of innovation look like that comprehensively integrates the normative dimensions of innovation? Is a concept of responsible innovation even logical or plausible from an economic perspective, or should innovation not always be responsible from a theoretical perspective? What might a theoretical differentiation between real innovation and "mere upgrades" (Soete 2013; Komlos 2016) look like? Why should not excessive innovation be acknowledged to be just as much a possibility as is excessive investment in scientific research, or in industrial R&D (David 2012)?

Responsible innovation and business ethics:

How can a stronger link between responsible innovation and business ethics be achieved both in theory and in practice? What is a comprehensive theory of responsible business innovation that combines the process-oriented perspective of responsible innovation with the governance-oriented and organisational perspective of business ethics? How can the practical implementation of responsible innovation in business be guided by more established practical approaches of business ethics?

Responsible innovation and acceptance research:

How can a scalable conceptualisation of various types of participation in the sense of responsible innovation be developed, which can replace a passive operationalisation of acceptance in the sense of approval/disapproval and support/resistance? How can the societal negotiation process on the costs and benefits of innovations and their societal distribution in the background of voiced acceptance (Juma 2016; Mokyr 2000) be related more directly to innovation?

Towards a field of innovation ethics:

Is it useful to specify a field of innovation ethics and what gap, if any, does it fill between neighbouring fields of research? On what philosophical basis could a comprehensive theory of innovation ethics be built upon? What could a normative theory look like that integrates all four levels of the innovation ethics circle (see 2.2.2)? Can a general framework for empirical research on the different levels be derived from this?

5.5 Implications for practice

The following chapter presents implications for key actors in the innovation system that can be developed on the basis of the results and discussion of this dissertation. Since a number of specific consequences and recommendations can be derived from the background of the individual articles which take up concrete results of the individual case studies, a few overarching implications for the prospects of responsible innovation will be highlighted in this section. The selection and systematisation of the actors of the innovation system is based on the so-called Quadruple Helix Model (Carayannis and Campbell 2012b) (see also section 2.1.3), according to which not only academia, industry, and policy are core stakeholders, but also actors from civil society.

5.5.1 Business

Firms are perhaps the most important actors when it comes to the development and, above all, the introduction and utilisation of innovations. They are often the ones who decide which innovations find their way to society and in what way. Accordingly, however, they often also play an important role when it comes to the negative effects of innovations (Giuliani 2018). Against the background of the results of this dissertation, the following recommendations can be derived for a better alignment in the sense of responsible innovation.

Businesses should develop a more systematic approach to identify, evaluate, and monitor the societal impacts of their (innovation) practices

There are a number of pertinent reasons why it is becoming increasingly important for companies to systematically identify and evaluate the (potential) impacts of their activities and to continuously reflect on and adapt strategies, products, and measures accordingly (Biggi and Giuliani 2021). The reasons include not least the complex and widespread effects of entrepreneurial practices against the backdrop of a digitalised and globalised economic system, the rising costs of responding late or not at all to (negative) effects, and the growing societal demand for responsible and responsive business behaviour. In order to take these developments adequately into account, a substantial shift in attention is required, for example through a significant increase in the internal capacities required for this, fundamentally changed organisational structures and processes, and a strategic opening for new networks and cooperations. For a first beginning, this can also include, for example, incorporating the diverse

knowledge stocks that arise in the context of social, environmental, or organisational reporting initiatives as part of knowledge generation in the sense of innovation (Mulgan 2016).

Businesses need to reflect and re-evaluate what they consider as innovation as well as their purpose for engaging in innovation

Precisely because companies assume a pivotal function in the innovation system, it is important that they do more justice to this role by making their concepts and goals for innovation transparent and by reflecting on them. This involves giving greater consideration to the first phase of the innovation ethics circle, i.e. analysing the framework conditions and prerequisites of one's own innovation activities. This is by no means a purely critical or backward-looking activity: rather, it is about discovering new and substantial fields for innovation that can make an effective positive contribution to entrepreneurial and societal progress. It is necessary to allocate a substantial share of the resources currently devoted to innovation in the sense of mere upgrades to activities that aim at "true innovation" instead (Komlos 2016; Soete 2019, 2013).

Business should start developing a new understanding of value-creation

Insofar as innovation is pursued by companies with the aim of creating value, they should develop a differentiated approach to value (creation) against the background of a comprehensive understanding as societal actors (Mazzucato 2018). In addition to the analysis and reflection of the existing, mostly implicit conceptualisation, this means a transparent identification of the relevant values of all stakeholders involved. On this basis, a decision, prioritisation, and assessment of the values to be realised can ultimately be made and corresponding strategies can be developed.

5.5.2 Policy

The political design of framework conditions plays a crucial role for innovation in general and the prospects of responsible innovation in particular. Innovation is determined to a considerable extent by incentives such as subsidies, by the legal containment of markets, or the development of norms and standards. Against the background of the results of this dissertation, the following recommendations can be derived for a stronger reorientation of innovation in the sense of responsible innovation.

Innovation policy should be based on a more comprehensive understanding of innovation, in particular by recognizing its political nature

As has been shown at several occasions throughout this dissertation, the general understanding of innovation today neglects the different normative dimensions and takes for granted that innovation is good per se (Coad et al. 2021). This holds especially true for the conceptualisation that is often implicitly used as a basis in policy-making (Schubert 2015a). Hence, understanding innovation first and foremost as a political concept means opening up the one-dimensional perspective on innovation as a scientific-technical and economic instrument for growth and understanding innovation as a socio-political field of policy. In this way, the complex normative interdependencies can be operationalised conceptually on a level with economic and other targeted effects. Moreover, the framing of innovation as a political construct means that disputes about the (non-)acceptance of certain technologies are part of the democratic process of societal progress through innovation – and not its external barrier. And finally, it also makes clear that innovation is one, but not the only arena for societal progress.

The public sector needs to be better equipped to foster responsible innovation

In view of the manifold societal implications of innovation, it becomes clear what a complex field innovation policy is. Against this background, it is important that the public sector not only disposes over the adequate resources and capacities, but is also able to coordinate and steer them in an inter-departmental manner (Soete 2013). For example, as a result of the conceptual neglect of the socio-ethical dimensions of innovation, the capacity to identify and assess the multiple potential impacts at an early stage as part of the innovation policy-making process has not received sufficient

attention. Such foresight and technology assessment capacities do not exist today in adequate relation to corresponding aspects such as the technical-economic stimulation of innovation. The lack of necessary competencies for a comprehensive approach to innovation is further exacerbated when the epistemic and temporal aspects of innovation are taken into account. Innovations themselves and their possible effects are often not plannable or recognisable in advance, and societal consequences often occur with a significant time delay. This is why activities such as long-term monitoring are becoming all the more important. Furthermore, opportunities for more comprehensive testing as well as simulation strategies for innovations, as already common in certain sectors such as pharmaceuticals, should be given greater priority.

Acknowledging the directionality of innovation also means to foster the societal dimension of adopting a mission-oriented approach to innovation policy

Although the directionality of innovation is in fact hardly a novelty (Nelson 1977), a mission-oriented orientation of innovation policy has recently come increasingly into focus (Mazzucato 2017, 2016, 2021). In addition to the European Commission, the Expert Commission for Research and Innovation (EFI), which is influential in the German context, called for a new mission orientation in research and innovation policy in its recent report (Expertenkomission Forschung und Innovation (EFI) 2021). Against the background of this dissertation, it should be emphasised that when formulating missions and goals, especially societal aspects should be addressed. In other words, the normative dimensions must be communicated and discussed in a transparent and balanced way. Precisely because missions may already appeal to societal goals and in this sense may seem to be taken for granted as common sense, it is crucial to identify and openly weigh hidden societal aspects and normative-ethical dimensions. Otherwise, such an approach runs the risk of falling prey to the dangerous misunderstanding that it is already "validated" from a normative-ethical point of view through the definition of a society-oriented mission. In contrast, a decidedly normative orientation of innovation policy increases the burden of legitimisation and normative-ethical dimensions gain additional importance.

5.5.3 Academia

In light of the above findings, implications can also be drawn for academia as an actor in the innovation system – beyond avenues for further research. For a stronger reorientation of innovation in the sense of responsible innovation, the following recommendations can thus be deduced:

Scientific policy advice in the field of innovation should follow a more holistic perspective

Particularly in the field of innovation, scientific policy advice plays a crucial role since by its very nature it often deals with new scientific developments and technologies. In line with the approach of responsible innovation, individual, often discipline-specific studies, forecasts, or assessments should be systematically complemented and combined with other scientific perspectives that address further aspects or dimensions of an issue. In other words, it seems important that scientific policy advice in the field of innovation is provided from a holistic perspective and, say, systematically complements economic modelling with relevant (normative) perspectives from other disciplines (Komlos 2016). This could be done, for example, through a common framework that covers different dimensions of innovation.

Innovation figures and statistics to measure innovation should be expanded beyond publications, patents, or GDP

Building on the first recommendation, central instruments of scientific policy advice in the field of innovation, namely key statistics and figures, should be enhanced and adapted. As individual figures such as the number of patents, publications or GDP growth, the expenditure on research and development, or labour market figures only ever allow a narrow view and limited conclusions to be drawn, a broader set of quantitative indicators should be developed that also attempt to cover broader societal aspects (Biggi and Giuliani 2021). Accordingly, the aim should be to base the general monitoring on a comprehensive understanding of innovation as a socio-political issue.

To support and engage in responsible innovation research organisations should strategically open up towards civil society

In the sense of the central tenet of responsible innovation, according to which societal perspectives and actors should be actively involved in innovation at an early stage, research organisations face the task of promoting inclusion (Schütz 2020). Given that this is already a widely researched subject, only two specific aspects should be pointed out here: First, in the sense of the preference learning aspect emphasised in the first article, research organisations should actively contribute not only to communicating new findings and technologies, but also to helping individuals form their own preferences about them – and thus focus on psychological and pedagogical levels in the sense of empowerment and political education. And secondly, participation is also about making new knowledge and technologies available to civil society actors. For applied research in particular, this means encouraging cooperation with civil society for the exploitation of new technologies to a similar extent as the transfer to industry. In addition to questions of intellectual property, patenting, licensing, and open source, this also concerns the substantial increase in the level of technological maturity in the public sector and civil society.

5.5.4 Civil society

A central element of responsible innovation is the early and strategic integration of societal values. To this end, civil society plays an important role, for example as an active stakeholder in the innovation process (Arnkil et al. 2010). The important role of civil society has recently been highlighted, especially in the context of the digital transformation (Beining, Bihr, and Heumann 2020; Rasmussen 2018). In the following, three overarching challenges for civil society in the context of responsible innovation will be presented against the backdrop of the findings of the dissertation.

Civil society needs to act as a seismograph for evaluating the societal costs and benefits of innovation

Civil society actors should develop a subtle sensorium for identifying, analysing, and comprehensively evaluating the ramifications of innovations from a societal perspective. This requires, on the one hand, necessary expertise and networks, and on the other hand, the initiation and implementation of societal negotiation processes as well as the consolidation of arguments and societal positions. To do so, the continuing support of media actors and a functioning journalistic system is needed (Buhmann and Fieseler 2021).

It should be not least due to a strong 'third sector' that innovation is understood as a socio-political topic

The goal of civil society engagement should not only be the critical accompaniment of specific developments but also to frame and position innovation as a socio-political issue. In this way, the normative-ethical dimensions are to be taken into account by adapting the definitional-conceptual framework of innovation.

The role of civil society in innovation is, among other things, to be the guardian of democracy and find the societal optimum between change and stability

The societal struggle around innovation which is commonly expressed in acceptance, dissemination, or use, is the negotiation process between change and technological change on the one hand and societal stability on the other (Juma 2016; Mokyr 2000). Civil society should not only be a mediator and medium of this democratic process

but should also make it visible as such. For insofar as too rapid and drastic change processes pay the price of social inequality and, conversely, societal uniformity does not mean progress and prosperity, it is the task of civil society to shape innovation in harmony with societal cohesion. In addition to dealing with this enormous challenge, it is first of all vital to recognise its urgency and relevance.

6. Conclusion

Over the past centuries, innovations have created prosperity and led to significant societal progress. Particularly in recent decades, technological change has accelerated, gained momentum, and increased the intensity of impact in terms of both quality and scope (Brynjolfsson and McAfee 2014). Think of, for example, advances in the field of digital technologies and their impact on almost all sectors of society, leading, amongst other things, to a „fourth industrial revolution" (Schwab 2016) and to fundamental changes in communication, business models, and how individuals as consumers or citizens interact with each other as well as with companies or public administration. But also in fields such as biotechnology or energy, breakthroughs such as new methods for genome editing or genetic vaccine development or hydrogen technologies are about to spark revolutionary changes. However, technological change can no longer be assumed to be synonymous with societal progress and prosperity. With increasing frequency and relevance, negative aspects of innovation have moved centre stage, be it in terms of detrimental health effects that have only been discovered with a longer time lag, rising costs due to damage to the environment and the climate, significant shifts in labour markets, or ethical implications with regard to surveillance and discrimination.

Given the importance of the societal dimensions and thus the normative side of innovation, it is not only the negative effects in the sense of unintended consequences or externalities that are put into focus, but rather the very purpose of innovation itself that is questioned. In spite of the fact that innovation is often understood, at least implicitly, as a "growth instrument", equated with progress, and understood as something per se good, the normative complexity of innovation comes to the fore. But „[i]f innovations are not an end in their own, what precisely are the ends they serve? What normative justification can be given for these ends" (Binder and Witt 2019, 343)? Even though R. Nelson noted as early as 1977 that the debate in this field "has been weak in setting forth normative criteria", no significant progress can be observed to date (Schubert 2015a). On the contrary, the simplistic assumption of innovation as something per se good seems to have become so entrenched over time that opening it up to more normative complexity appears ever more difficult. In this light, efforts to promote *responsible innovation* should be understood as a contribution to making the normative complexity of innovation visible and enabling a realignment with societal values. This is where this dissertation starts and situates responsible innovation against the background of the normative complexity of the concept itself.

Three individual case studies show that in order to deal with innovation and its normative dimensions in an appropriate and effective way, responsible innovation should include a perspective from economic theory, address the societal role of companies, and adapt the conceptualisation of the term "acceptance". Responsible innovation can thereby contribute to a re-politicisation of innovation, so that innovation is (again) understood as a societal project. Through technological change, society creates itself anew each time, decides which futures it envisages, and which distributions of costs and benefits, opportunities and risks, are fair and desirable for society. Innovation is hence a pivotal arena for addressing the values that should determine societal development and progress. Innovation as a political concept is not (only) value creation in the financial sense (Mazzucato 2018), but should reflect and promote all societal values and their prioritisation. It then becomes clear that the relationship between innovation and normativity does not only consist of negative societal effects on the one hand or instrumental opportunities to cope with societal challenges on the other. Innovation is in itself a field in which societal negotiation processes take place, social inequalities are discussed, and societal cohesion is shaped. In other words: Innovation is a place of democracy – and responsible innovation an approach to shaping it. That innovation is able to assume this task is of major importance – not least in view of growing societal inequalities, increased political polarisation, and trends that even threaten democracy (Przeworski 2019; Schäfer and Zürn 2021). Against this background, this dissertation presents ways of re-politicising the concept through responsible innovation.

References

Abbott, Andrew. 1983. 'Professional Ethics'. *American Journal of Sociology* 88 (5): 583–96. https://doi.org/https://doi.org/10.1086/227762.

Abels, Gabriele, and Alfons Bora. 2004. *Demokratische Technikbewertung*. Bielefeld: Transcript.

Abolhasani, Zahra, Alireza Hassanzadeh, S. Sepehr Ghazinoory, and Aliasghar Pourezzat. 2014. 'A Justice-Oriented Innovation System: A Grounded Theory Approach'. *Social Justice Research* 27 (3): 369–94. https://doi.org/10.1007/s11211-014-0218-2.

Achterberg, Peter, Dick Houtman, Samira Van Bohemen, and Katerina Manevska. 2010. 'Unknowing but Supportive? Predispositions, Knowledge, and Support for Hydrogen Technology in the Netherlands'. *International Journal of Hydrogen Energy* 35 (12): 6075–83. https://doi.org/10.1016/j.ijhydene.2010.03.091.

Acs, Zoltan J., David B. Audretsch, Pontus Braunerhjelm, and Bo Carlsson. 2012. 'Growth and Entrepreneurship'. *Small Business Economics* 39 (2): 289–300. https://doi.org/10.1007/s11187-010-9307-2.

Adams, Richard, Sally Jeanrenaud, John Bessant, David Denyer, and Patrick Overy. 2016. 'Sustainability-Oriented Innovation: A Systematic Review'. *International Journal of Management Reviews* 18 (2): 180–205. https://doi.org/10.1111/ijmr.12068.

Adner, Ron, and Rahul Kapoor. 2010. 'Value Creation in Innovation Ecosystems: How the Structure of Technological Interdependence Affects Firm Performance in New Technology Generations'. *Strategic Management Journal* 31 (3): 306–33. https://doi.org/10.1002/smj.821.

Aghion, Philippe, and Peter Howitt. 1992. 'A Model of Growth Through Creative Destruction'. *Econometrica* 60 (2): 323. https://doi.org/10.2307/2951599.

AI Ethics Impact Group. 2020. 'From Principles to Practice: An Interdisciplinary Framework to Operationalise AI Ethics'. *Bertelsmann Stiftung*. Gütersloh. https://doi.org/10.11586/2020013.

AI4People. 2018. 'The AI4People's Ethical Framework For A Good AI Society: Opportunities, Risks, Principles, And Recommendations'. Brussels.

Aitken, Mhairi. 2010. 'Why We Still Don't Understand the Social Aspects of Wind Power: A Critique of Key Assumptions within the Literature'. *Energy Policy* 38 (4): 1834–41. https://doi.org/10.1016/j.enpol.2009.11.060.

Aizenberg, Evgeni, and Jeroen van den Hoven. 2020. 'Designing for Human Rights in AI'. *Big Data & Society* 7 (2): 1–14. https://doi.org/10.1177/2053951720949566.

Akerlof, George A., and Robert J. Shiller. 2015. *Phishing for Phools: The Economics of Manipulation and Deception*.

Altenburg, Tilman. 2009. 'Building Inclusive Innovation Systems in Developing Countries: Challenges for IS Research'. In *Handbook of Innovation Systems and Developing Countries*, 33–56. Edward Elgar Publishing. https://doi.org/10.4337/9781849803427.00008.

Altmann, M., and C. Graesel. 1998. 'The Acceptance of Hydrogen Technologies'. Germany.

Andreas Jørgensen, Agathe Le Bocq, Liudmila Nazarkina, and Michael Hauschild. 2008. 'Methodologies for Social Life Cycle Assessment'. *International Journal of Life Cycle Assessment* 2 (2): 96–103. http://www.me.mtu.edu/~jwsuther/colloquium/Nesbitt_Methodolgies for SLCA.pdf.

Arnkil, Robert, Anu Järvensivu, Pasi Koski, and Tatu Piirainen. 2010. *Exploring Quadruple Helix: Outlining User-Oriented Innovation Models. Final Report on Quadruple Helix Research for the CLIQ Project. Tampere: The CLIQ*. http://www.arnkildialogues.com/files/testataan.kotisivukone.com/julkaisut/exploring_quadruple_helix-2010-1.pdf.

Arntz, Melanie, Terry Gregory, and Ulrich Zierahn. 2017. 'Revisiting the Risk of Automation'. *Economics Letters* 159: 157–60. https://doi.org/10.1016/j.econlet.2017.07.001.

Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, et al. 2019. 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI'. http://arxiv.org/abs/1910.10045.

Ashworth, P., K. Witt, M. Ferguson, and S. Sehic. 2019. 'Developing Community Trust in Hydrogen'. Brisbane. http://www.coagenergycouncil.gov.au/sites/prod.energycouncil/files/publications/documents/nhs-developing-community-trust-in-hydrogen-report-2019_0.pdf.

Ast, Federico. 2019. 'The Deliberative Test, a New Procedural Method for Ethical Decision Making in Integrative Social Contracts Theory'. *Journal of Business Ethics* 155 (1): 207–21. https://doi.org/10.1007/s10551-017-3481-y.

Asveld, Lotte, Rietje van Dam-Mieras, Tsjalling Swierstra, Saskia Lavrijssen, Kees Linse, and Jeroen van den Hoven, eds. 2017. *Responsible Innovation 3*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-64834-7.

Asveld, Lotte, Jurgen Ganzevles, and Patricia Osseweijer. 2015. 'Trustworthiness and Responsible Research and Innovation: The Case of the Bio-Economy'. *Journal of Agricultural and Environmental Ethics* 28 (3): 571–88. https://doi.org/10.1007/s10806-015-9542-2.

Auer, Alexander, and Katharina Jarmai. 2017. 'Implementing Responsible Research and Innovation Practices in SMEs: Insights into Drivers and Barriers from the Austrian Medical Device Sector'. *Sustainability (Switzerland)* 10 (1): 1–18. https://doi.org/10.3390/su10010017.

Australian Public Service. 2020. 'The Australian Public Service Framework for Engagement and Participation'.

Autio, E., and L.D.W. Thomas. 2014. 'Innovation Ecosystems: Implications for Innovation Management?' In *The Oxford Handbook of Innovation Management*, edited by M. Dodgson, D. M. Gann, and N. Phillips, 204–28. Oxford: Oxford University Press.

Autor, David H. 2015. 'Why Are There Still So Many Jobs? The History and Future of Workplace Automation'. *Journal of Economic Perspectives* 29 (3): 3–30. https://doi.org/10.1257/jep.29.3.3.

Bacon, Francis. 1625. *Of Innovations*.

Bakker, Erik de, Carolien de Lauwere, Anne Charlotte Hoes, and Volkert Beekman. 2014. 'Responsible Research and Innovation in Miniature: Information Asymmetries Hindering a More Inclusive "nanofood" Development'. *Science and Public Policy* 41 (3): 294–305. https://doi.org/10.1093/scipol/scu033.

Banta, David. 2009. 'What Is Technology Assessment?' *International Journal of Technology Assessment in Health Care* 25 (SUPPL.S1): 7–9. https://doi.org/10.1017/S0266462309090333.

Barben, Daniel, Erik Fisher, Cynthia Selin, and David H. Guston. 2008. 'Anticipatory Governance of Nanotechnology: Foresight, Engagement, and Integration'. In *The Handbook of Science and Technology Studies*, edited by Edward J. Hackett, Olga Amsterdamska, Michael Lynch, and Judy Wajcman, 979–94. Cambridge, Massachusetts; London, England: The MIT Press.

Barocas, Solon, Moritz Hardt, and Arvind Narayanan. 2019. *Fairness and Machine Learning*. fairmlbook.org. http://www.fairmlbook.org.

Bartneck, Christoph, Christoph Lütge, Alan Wagner, and Sean Welsh. 2021. 'Risks in the Business of AI'. In , 45–53. https://doi.org/10.1007/978-3-030-51110-4_6.

Batel, Susana. 2018. 'A Critical Discussion of Research on the Social Acceptance of Renewable Energy Generation and Associated Infrastructures and an Agenda for the Future'. *Journal of Environmental Policy & Planning* 20 (3): 356–69. https://doi.org/10.1080/1523908X.2017.1417120.

Batel, Susana, Patrick Devine-Wright, and Torvald Tangeland. 2013. 'Social Acceptance of Low Carbon Energy and Associated Infrastructures: A Critical Discussion'. *Energy*

Policy 58: 1–5. https://doi.org/10.1016/j.enpol.2013.03.018.

Baumol, William J. 1990. 'Entrepreneurship : Productive, Unproductive, and Destructive'. *Journal of Political Economy* 98 (5): 893–921. http://www.jstor.org/stable/2937617.

Bausinger, Tobias, Eric Bonnaire, and Johannes Preuß. 2007. 'Exposure Assessment of a Burning Ground for Chemical Ammunition on the Great War Battlefields of Verdun'. *Science of The Total Environment* 382 (2–3): 259–71. https://doi.org/10.1016/j.scitotenv.2007.04.029.

Baxter, Jamie. 2017. 'Energy Justice: Participation Promotes Acceptance'. *Nature Energy* 2 (8). https://doi.org/10.1038/nenergy.2017.128.

Bechmann, Gotthard, Michael Decker, Ulrich Fiedeler, and Bettina Johanna Krings. 2007. 'Technology Assessment in a Complex World'. *International Journal of Foresight and Innovation Policy* 3 (1): 6. https://doi.org/10.1504/IJFIP.2007.011419.

Beck, Ulrich. 1986. *Risikogesellschaft. Auf Dem Weg in Eine Andere Moderne*. Frankfurt am Main: Suhrkamp Verlag.

Beining, Leonie, Peter Bihr, and Stefan Heumann. 2020. 'Towards a European AI & Society Ecosystem Why We Need It and How to Empower It'. Berlin.

Bell, Derek, Tim Gray, and Claire Haggett. 2005. 'The "Social Gap" in Wind Farm Siting Decisions: Explanations and Policy Responses'. *Environmental Politics* 14 (4): 460–77. https://doi.org/10.1080/09644010500175833.

Bell, Derek, Tim Gray, Claire Haggett, and Joanne Swaffield. 2013. 'Re-Visiting the "Social Gap": Public Opinion and Relations of Power in the Local Politics of Wind Energy'. *Environmental Politics* 22 (1): 115–35. https://doi.org/10.1080/09644016.2013.755793.

Benaich, Nathan, and Nathan Benaich. 2019. 'State of AI Report'. London, United Kingdom.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. 'On the Dangers of Stochastic Parrots'. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. New York, NY, USA: ACM. https://doi.org/10.1145/3442188.3445922.

Benkler, Yochai. 2019. 'Don't Let Industry Write the Rules for AI'. *Nature* 569 (7755): 161–161. https://doi.org/10.1038/d41586-019-01413-1.

Benoît, Catherine, Gregory A. Norris, Sonia Valdivia, Andreas Ciroth, Asa Moberg, Ulrike Bos, Siddharth Prakash, Cassia Ugaya, and Tabea Beck. 2010. 'The Guidelines for Social Life Cycle Assessment of Products: Just in Time!' *International Journal of Life Cycle Assessment* 15 (2): 156–63. https://doi.org/10.1007/s11367-009-0147-8.

Bernard, H. Russell. 2018. *Research Methods in Anthropology. Qualitative and Quantitative*

Approaches. 6th Editio. Lanham, Boulder, New York, Toronto, and Plymouth, UK: Rowman & Littlefield Publishers, Inc.

Bertelsmann Stiftung. 2018. 'Ethik Für Algorithmiker: Was Wir von Erfolgreichen Professionsethiken Lernen Können'. Gütersloh. https://doi.org/10.11586/2018033.

Bertsimas, Dimitris, Arthur Delarue, and Sebastien Martin. 2019. 'Optimizing Schools' Start Time and Bus Routes'. *Proceedings of the National Academy of Sciences of the United States of America* 116 (13): 5943–48. https://doi.org/10.1073/pnas.1811462116.

Bhaskar, Abhinav, Mohsen Assadi, and Homam Nikpey Somehsaraei. 2020. 'Decarbonization of the Iron and Steel Industry with Direct Reduction of Iron Ore with Green Hydrogen'. *Energies* 13 (3): 758. https://doi.org/10.3390/en13030758.

Biggi, Gianluca, and Elisa Giuliani. 2021. 'The Noxious Consequences of Innovation: What Do We Know?' *Industry and Innovation* 28 (1): 19–41. https://doi.org/10.1080/13662716.2020.1726729.

Bijker, Wiebe E., Thomas P. Hughes, and Trevor Pinch, eds. 1987. *The Social Construction of Technological Systems : New Directions in the Sociology and History of Technology*. Cambridge, Massachusetts: MIT Press.

Bijker, Wiebe E., and John Law, eds. 1992. *Shaping Technology / Building Society*. Cambridge, Massachusetts: MIT Press.

Bimber, Bruce A. 1996. *The Politics of Expertise in Congress : The Rise and Fall of the Office of Technology Assessment*. New York: State University of New York Press.

Binder, Martin. 2013. 'Innovativeness and Subjective Well-Being'. *Social Indicators Research* 111 (2): 561–78. https://doi.org/10.1007/s11205-012-0020-1.

Binder, Martin, and Ulrich Witt. 2011. 'As Innovations Drive Economic Growth, Do They Also Raise Well-Being?' *Papers on Economics and Evolution, No. 1105*. http://nbn-resolving.de/urn:nbn:de:gbv:27-20110630-114932-7.

——. 2019. 'As Innovations Drive Economic Change, Do They Also Improve Our Welfare?' In *Understanding Economic Change*, edited by Ulrich Witt and Andreas Chai, 343–68. Cambridge & New York: Cambridge University Press. https://doi.org/10.1017/9781316477168.011.

Binns, Reuben. 2018. 'Algorithmic Accountability and Public Reason'. *Philosophy and Technology* 31 (4): 543–56. https://doi.org/10.1007/s13347-017-0263-5.

——. 2020. 'On the Apparent Conflict between Individual and Group Fairness'. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 514–24. New York, NY, USA: ACM. https://doi.org/10.1145/3351095.3372864.

Bland, J M., and D. G Altman. 1995. 'Statistics Notes: Multiple Significance Tests: The Bonferroni Method'. *BMJ* 310 (6973): 170–170. https://doi.org/10.1136/bmj.310.6973.170.

Blok, V., L. Hoffmans, and E.F.M. Wubben. 2015. 'Stakeholder Engagement for Responsible Innovation in the Private Sector: Critical Issues and Management Practices'. *Journal on Chain and Network Science* 15 (2): 147–64. https://doi.org/10.3920/JCNS2015.x003.

Blok, Vincent. 2018a. 'From Participation to Interruption: Toward an Ethics of Stakeholder Engagement, Participation and Partnership in CSR and Responsible Innovation'. In *Handbook Responsible Innovation: A Global Resource*, edited by René Von Schomberg and J. Hankins, 1–22. Cheltenham, United Kingdom: Edward Elgar Publishing.

——. 2018b. 'Philosophy of Innovation: A Research Agenda'. *Philosophy of Management* 17 (1): 1–5. https://doi.org/10.1007/s40926-017-0080-z.

——. 2019a. 'From Participation to Interruption: Toward an Ethics of Stakeholder Engagement, Participation and Partnership in Corporate Social Responsibility and Responsible Innovation'. In *International Handbook on Responsible Innovation*, 243–57. Edward Elgar Publishing. https://doi.org/10.4337/9781784718862.00024.

——. 2019b. 'Innovation as Ethos'. In *Handbook of Philosophy of Management*, edited by C. Neesham and S. Segal, 1–14. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-48352-8_19-1.

——. 2020. 'Towards an Ontology of Innovation: On the New, the Political-Economic Dimension and the Intrinsic Risks Involved in Innovation Processes'. In *Routledge Handbook of Philosophy of Engineering*, edited by Diane P. Michelfelder and Neelke Doorn. New York: Routledge. https://www.taylorfrancis.com/books/9781351996563.

Blok, Vincent, and Pieter Lemmens. 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 *Responsible Innovation 2*, 19–35. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17308-5_2.

Bloor, Michael, Jane Frankland, Michelle Thomas, and Kate Robson. 2001. *Focus Groups in Social Research*. London: SAGE Publications Ltd.

Boatright, John R. 2000. 'Contract Theory and Business Ethics: A Review of Ties That Bind'. *Business and Society Review* 105 (4): 452–66. https://doi.org/10.1111/0045-3609.00094.

Boddington, Paula. 2017. *Towards a Code of Ethics for Artificial Intelligence*. Artificial Intelligence: Foundations, Theory, and Algorithms. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-60648-4.

Boenink, Marianne. 2013. 'The Multiple Practices of Doing "Ethics in the Laboratory": A Mid-Level Perspective'. In *Ethics on the Laboratory Floor*, 57–78. London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137002938_4.

Bögel, Paula, Christian Oltra, Roser Sala, Monica Lores, Paul Upham, Elisabeth Dütschke, Uta Schneider, and Paul Wiemann. 2018. 'The Role of Attitudes in Technology Acceptance Management: Reflections on the Case of Hydrogen Fuel Cells in Europe'. *Journal of Cleaner Production* 188: 125–35. https://doi.org/10.1016/j.jclepro.2018.03.266.

Bonin, Holger, Terry Gregory, and Ulrich Zieran. 2015. *Ubertragung Der Studie von Frey/Osborne (2013) Auf Deutschland*. https://doi.org/ISSN 0174-4992.

Bonnemains, Vincent, Claire Saurel, and Catherine Tessier. 2018. 'Embedded Ethics: Some Technical and Ethical Challenges'. *Ethics and Information Technology* 20 (1): 41–58. https://doi.org/10.1007/s10676-018-9444-x.

Bontems, Vincent Karim. 2014. 'What Does Innovation Stand for? Review of a Watchword in Research Policies'. *Journal of Innovation Economics* 15 (3): 39. https://doi.org/10.3917/jie.015.0039.

Boon, Wouter, and Jakob Edler. 2018. 'Demand, Challenges, and Innovation. Making Sense of New Trends in Innovation Policy'. *Science and Public Policy* 45 (February): 435–47. https://doi.org/10.1093/scipol/scy014.

Borgesius, Frederik J.Zuiderveen, Judith Möller, Sanne Kruikemeier, Ronan Fathaigh, Kristina Irion, Tom Dobber, Balazs Bodo, and Claes de Vreese. 2018. 'Online Political Microtargeting: Promises and Threats for Democracy'. *Utrecht Law Review* 14 (1): 82–96. https://doi.org/10.18352/ulr.420.

Bosch, Stephan, and Matthias Schmidt. 2020. 'Wonderland of Technology? How Energy Landscapes Reveal Inequalities and Injustices of the German Energiewende'. *Energy Research & Social Science* 70 (December): 101733. https://doi.org/10.1016/j.erss.2020.101733.

Boudet, Hilary S. 2019. 'Public Perceptions of and Responses to New Energy Technologies'. *Nature Energy* 4 (6): 446–55. https://doi.org/10.1038/s41560-019-0399-x.

Brand, Teunis. 2020. 'Food for Deliberation: Philosophical Reflections on Responsible Innovation in the Business Context'. Wageningen University. https://doi.org/10.18174/509524.

Brand, Teunis, and Vincent Blok. 2019. 'Responsible Innovation in Business: A Critical Reflection on Deliberative Engagement as a Central Governance Mechanism'. *Journal of Responsible Innovation* 6 (1): 4–24. https://doi.org/10.1080/23299460.2019.1575681.

Brand, Teunis, Vincent Blok, and Marcel Verweij. 2020. 'Stakeholder Dialogue as Agonistic Deliberation: Exploring the Role of Conflict and Self-Interest in Business-NGO Interaction'. *Business Ethics Quarterly* 30 (1): 3–30. https://doi.org/10.1017/beq.2019.21.

Braun, Robert, and Erich Griessler. 2018. 'More Democratic Research and Innovation'. *Journal of Science Communication* 17 (03): 1–7. https://doi.org/10.22323/2.17030304.

Brennan, Geoffrey, and James M. Buchanan. 1985. *The Reason of Rules: Constitutional Political Economy*. Cambridge: Cambridge University Press.

Brennan, Noreen, and Thomas M Van Rensburg. 2016. 'Wind Farm Externalities and Public Preferences for Community Consultation in Ireland: A Discrete Choice Experiments Approach'. *Energy Policy* 94 (July): 355–65. https://doi.org/10.1016/j.enpol.2016.04.031.

Breukers, Sylvia, and Maarten Wolsink. 2007. 'Wind Power Implementation in Changing Institutional Landscapes: An International Comparison'. *Energy Policy* 35 (5): 2737–50. https://doi.org/10.1016/j.enpol.2006.12.004.

Brudermüller, Martin, Reiner Hoffmann, Henning Kagermann, Reimund Neugebauer, and Günther Schuh. 2021. 'Innovations for a European Green Deal'. Munich.

Bruijnis, M. R.N., V. Blok, E. N. Stassen, and H. G.J. Gremmen. 2015. 'Moral "Lock-In" in Responsible Innovation: The Ethical and Social Aspects of Killing Day-Old Chicks and Its Alternatives'. *Journal of Agricultural and Environmental Ethics* 28 (5): 939–60. https://doi.org/10.1007/s10806-015-9566-7.

Brusoni, Stefano, and Antonino Vaccaro. 2017. 'Ethics, Technology and Organizational Innovation'. *Journal of Business Ethics* 143 (2): 223–26. https://doi.org/10.1007/s10551-016-3061-6.

Bryden, John, and Stig S. Gezelius. 2017. 'Innovation as If People Mattered: The Ethics of Innovation for Sustainable Development'. *Innovation and Development* 7 (1): 101–18. https://doi.org/10.1080/2157930X.2017.1281208.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton & Company.

Buchanan, Allen, Tony Cole, and Robert O. Keohane. 2011. 'Justice in the Diffusion of Innovation'. *Journal of Political Philosophy* 19 (3): 306–32. https://doi.org/10.1111/j.1467-9760.2009.00348.x.

Buchanan, James M. 1975. *The Limits of Liberty. Between Anarchy and Leviathan*. Chicago: Chicago University Press.

Buenstorf, Guido, Uwe Cantner, Horst Hanusch, Michael Hutter, Hans-Walter Lorenz, and Fritz Rahmeyer. 2013a. 'Editorial: The Two Sides of Innovation'. In *The Two Sides of Innovation*, 1–7. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-01496-8_1.

——. , eds. 2013b. *The Two Sides of Innovation*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-01496-8.

Bughin, Jacques, James Manyika, Jonathan Woetzel, Frank Michael Mattern, Susan Chui, Anu Lund, Sree Madgavkar, et al. 2017. 'A Future That Works: Automation, Employment, and Productivity'. *McKinsey Global Institute*, no. January: 1–28. http://njit2.mrooms.net/pluginfile.php/688844/mod_resource/content/1/Executive Summary of McKinsey Report on Automation.pdf.

Buhmann, Alexander, and Christian Fieseler. 2021. 'Towards a Deliberative Framework for Responsible Innovation in Artificial Intelligence'. *Technology in Society* 64 (January): 101475. https://doi.org/10.1016/j.techsoc.2020.101475.

Buhmann, Alexander, Johannes Paßmann, and Christian Fieseler. 2020. 'Managing Algorithmic Accountability: Balancing Reputational Concerns, Engagement Strategies, and the Potential of Rational Discourse'. *Journal of Business Ethics* 163 (2): 265–80. https://doi.org/10.1007/s10551-019-04226-4.

Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit (BMU). 2019. 'Umweltbewusstsein in Deutschland 2018: Ergebnisse Einer Repräsentativen Bevölkerungsumfrage'. Berlin.

Bundesregierung der Bundesrepublik Deutschland. 2010. 'Energiekonzept Für Eine Umweltschonende, Zuverlässige Und Bezahlbare Energieversorgung'. Berlin.

——. 2020. 'Die Nationale Wasserstoffstrategie'. Berlin. https://www.bmwi.de/Redaktion/DE/Publikationen/Energie/die-nationale-wasserstoffstrategie.html.

Buranyi, Stephen. 2018. '"Dehumanising, Impenetrable, Frustrating": The Grim Reality of Job Hunting in the Age of AI'. Guardian. 2018. https://www.theguardian.com/inequality/2018/mar/04/dehumanising-impenetrable-frustrating-the-grim-reality-of-job-hunting-in-the-age-of-ai.

Buren III, Harry J. Van. 2001. 'If Fairness Is the Problem, Is Consent the Solution? Integrating ISCT and Stakeholder Theory'. Edited by Denis G. Arnold. *Business Ethics Quarterly* 11 (3): 481–99. https://doi.org/10.5840/10.2307/3857850.

Burg, Ryan. 2009. 'Deliberative Business Ethics'. *Journal of Business Ethics* 88 (S4): 665–83. https://doi.org/10.1007/s10551-009-0332-5.

Burget, Mirjam, Emanuele Bardone, and Margus Pedaste. 2017. 'Definitions and Conceptual Dimensions of Responsible Research and Innovation: A Literature Review'. *Science and Engineering Ethics* 23 (1): 1–19. https://doi.org/10.1007/s11948-016-9782-1.

Bush, Vannevar. 1945. 'Science the Endless Frontier. A Report to the President (on a Program for Postwar Scientific Research)'. Washington, D.C.

Byrne, Edmund F. 2018. 'Making Drones to Kill Civilians: Is It Ethical?' *Journal of Business Ethics* 147 (1): 81–93. https://doi.org/10.1007/s10551-015-2950-4.

Callon, Michel. 2009. 'Civilizing Markets: Carbon Trading between in Vitro and in Vivo Experiments'. *Accounting, Organizations and Society* 34: 535–548.

Calo, Ryan. 2018. 'Artificial Intelligence Policy: A Primer and Roadmap'. *U.C.D. L. Rev.* 51: 399.

Calton, Jerry M. 2006. 'Social Contracting in a Pluralist Process of Moral Sense Making: A Dialogic Twist on the ISCT'. *Journal of Business Ethics* 68 (3): 329–46. https://doi.org/10.1007/s10551-006-9017-5.

Carayannis, Elias G., and David F. J. Campbell. 2012a. *Mode 3 Knowledge Production in Quadruple Helix Innovation Systems. Mode 3 Knowledge Production in Quadruple Helix Innovation Systems*. https://doi.org/10.1007/978-1-4614-2062-0_1.

Carayannis, Elias G., and David F.J. Campbell. 2012b. *Mode 3 Knowledge Production in Quadruple Helix Innovation Systems*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-2062-0.

Carayannis, Elias G., and Ruslan Rakhmatullin. 2014. 'The Quadruple/Quintuple Innovation Helixes and Smart Specialisation Strategies for Sustainable and Inclusive Growth in Europe and Beyond'. *Journal of the Knowledge Economy* 5 (2): 212–39. https://doi.org/10.1007/s13132-014-0185-8.

Carayannis, Elias G, Thorsten D Barth, and David FJ Campbell. 2012. 'The Quintuple Helix Innovation Model: Global Warming as a Challenge and Driver for Innovation'. *Journal of Innovation and Entrepreneurship* 1 (1): 2. https://doi.org/10.1186/2192-5372-1-2.

Ceballos, Gerardo, Paul R. Ehrlich, and Rodolfo Dirzo. 2017. 'Biological Annihilation via the Ongoing Sixth Mass Extinction Signaled by Vertebrate Population Losses and Declines'. *Proceedings of the National Academy of Sciences* 114 (30): E6089–96. https://doi.org/10.1073/pnas.1704949114.

Chataway, Joanna, Rebecca Hanlin, and Raphael Kaplinsky. 2014. 'Inclusive Innovation: An Architecture for Policy Development'. *Innovation and Development* 4 (1): 33–54. https://doi.org/10.1080/2157930X.2013.876800.

Chatfield, Kate, Konstantinos Iatridis, Bernd C. Stahl, and Nearchos Paspallis. 2017.

'Innovating Responsibly in ICT for Ageing: Drivers, Obstacles and Implementation'. *Sustainability (Switzerland)* 9 (6): 1–22. https://doi.org/10.3390/su9060971.

Chen, Al Y.S., Roby B. Sawyers, and Paul F. Williams. 1997. 'Reinforcing Ethical Decision Making through Corporate Culture'. *Journal of Business Ethics* 16 (8): 855–65. https://doi.org/10.1023/A:1017953517947.

Chesbrough, Henry, Wim Vanhaverbeke, and Joel West, eds. 2006. *Open Innovation. Researching a New Paradigm*. Oxford.

Choi, Nia, and Satyajit Majumdar. 2015. 'Social Innovation: Towards a Conceptualisation'. In *Technology and Innovation for Social Change*, 7–34. New Delhi: Springer India. https://doi.org/10.1007/978-81-322-2071-8_2.

Chouldechova, Alexandra, and Aaron Roth. 2018. 'The Frontiers of Fairness in Machine Learning', October. http://arxiv.org/abs/1810.08810.

Christensen, Malene Vinther, Mika Nieminen, Marlene Altenhofer, Elise Tangcoigne, Niels Mejlgaard, Erich Griessler, and Adolf Filacek. 2020. 'What's in a Name? Perceptions and Promotion of Responsible Research and Innovation Practices across Europe'. *Science and Public Policy*, October. https://doi.org/10.1093/scipol/scaa040.

Ciampi, Francesco, Stefano Demi, Alessandro Magrini, Giacomo Marzi, and Armando Papa. 2021. 'Exploring the Impact of Big Data Analytics Capabilities on Business Model Innovation: The Mediating Role of Entrepreneurial Orientation'. *Journal of Business Research* 123 (February): 1–13. https://doi.org/10.1016/j.jbusres.2020.09.023.

Clausen, Jens, and Thomas Loew. 2009. 'CSR Und Innovation: Literaturstudie Und Befragung'. Berlin, Münster. http://www.4sustainability.org/downloads/Clausen-Loew_CSR-und-Innovation-LiteraturstudieundBefragung.pdf.

Coad, Alex, Paul Nightingale, Jack Stilgoe, and Antonio Vezzani. 2021. 'Editorial: The Dark Side of Innovation'. *Industry and Innovation* 28 (1): 102–12. https://doi.org/10.1080/13662716.2020.1818555.

Coase, Ronald H. 1960. 'The Problem of Social Cost'. *Journal of Law and Economics* 3: 1–44. http://www.jstor.org/stable/724810.

Coates, Joseph F. 1982. 'What Is Technology Assessment?' *Impact Assessment* 1 (1): 20–24. https://doi.org/10.1080/07349165.1982.9725925.

——. 2001. 'A 21st Century Agenda for Technology Assessment'. *Technological Forecasting and Social Change* 67 (2–3): 303–8. https://doi.org/10.1016/S0040-1625(01)00135-4.

Coeckelbergh, Mark. 2019. 'Ethics of Artificial Intelligence: Some Ethical Issues and Regulatory Challenges'. *Technology and Regulation*, 31–34. https://doi.org/10.26116/techreg.2019.003.

Cohen, Jacob. 1992. 'A Power Primer'. *Psychological Bulletin* 112 (1): 155–59. https://doi.org/10.1037/0033-2909.112.1.155.

Cooper, Paul. 2018. 'Https://Twitter.Com/PaulMMCooper/Status/989100350044082176'.

Cooper, Robert G. 1983. 'A Process Model for Industrial New Product Development'. *IEEE Transactions on Engineering Management* EM-30 (1): 2–11. https://doi.org/10.1109/TEM.1983.6448637.

——. 1986. 'An Investigation into the New Product Process: Steps, Deficiencies, and Impact'. *Journal of Product Innovation Management* 3 (2): 71–85. https://doi.org/10.1016/0737-6782(86)90030-5.

Cooper, Robert G. 1990. 'Stage-Gate Systems: A New Tool for Managing New Products'. *Business Horizons* 33 (3): 44–54. https://doi.org/10.1016/0007-6813(90)90040-I.

Corbett-Davies, Sam, and Sharad Goel. 2018. 'The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning', July. http://arxiv.org/abs/1808.00023.

Corbett-Davies, Sam, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. 'Algorithmic Decision Making and the Cost of Fairness'. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797–806. New York, NY, USA: ACM. https://doi.org/10.1145/3097983.3098095.

Corona, Blanca, Kossara P. Bozhilova-Kisheva, Stig I. Olsen, and Guillermo San Miguel. 2017. 'Social Life Cycle Assessment of a Concentrated Solar Power Plant in Spain: A Methodological Proposal'. *Journal of Industrial Ecology* 21 (6): 1566–77. https://doi.org/10.1111/jiec.12541.

Correljé, Aad, Eefje Cuppen, Marloes Dignum, Udo Pesch, and Behnam Taebi. 2015. 'Responsible Innovation in Energy Projects: Values in the Design of Technologies, Institutions and Stakeholder Interactions'. In *Responsible Innovation 2*, 183–200. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17308-5_10.

Cotton, Matthew. 2014. *Ethics and Technology Assessment: A Participatory Approach*. Vol. 13. Studies in Applied Philosophy, Epistemology and Rational Ethics. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-45088-4.

Cowls, Josh, Andreas Tsamados, Mariarosaria Taddeo, and Luciano Floridi. 2021. 'A Definition, Benchmark and Database of AI for Social Good Initiatives'. *Nature Machine Intelligence* 3 (February): 111–15. https://doi.org/10.1038/s42256-021-00296-0.

Cressey, Daniel. 2016. 'Bottles, Bags, Ropes and Toothbrushes: The Struggle to Track Ocean Plastics'. *Nature* 536 (7616): 263–65. https://doi.org/10.1038/536263a.

Crockford, Kade, and Joi Ito. 2017. 'Don't Blame the Algorithm for Doing What Boston

School Officials Asked'. The Boston Globe. 2017. https://www3.bostonglobe.com/opinion/2017/12/22/don-blame-algorithm-for-doing-what-boston-school-officials-asked/lAsWv1Rfwqmq6Jfm5ypLmJ/story.html?arc404=true.

d'Aquin, Mathieu, Pinelopi Troullinou, Noel E. O'Connor, Aindrias Cullen, Gráinne Faller, and Louise Holden. 2018. 'Towards an "Ethics by Design" Methodology for AI Research Projects'. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society - AIES '18*, 54–59. New York, New York, USA: ACM Press. https://doi.org/10.1145/3278721.3278765.

D'Aquin, Mathieu, Pinelopi Troullinou, Noel E. O'Connor, Aindrias Cullen, Gráinne Faller, and Louise Holden. 2018. 'Towards an "Ethics by Design" Methodology for AI Research Projects'. *AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 54–59. https://doi.org/10.1145/3278721.3278765.

Daniels, Norman, and James E. Sabin. 2002. *Setting Limits Fairly: Can We Learn to Share Medical Resources?* Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195149364.001.0001.

Datenethikkomission. 2019. 'Gutachten Der Datenethikkommission'. Berlin.

David, Paul A. 2012. 'The Innovation Fetish among the Economoi'. In *The Rate and Direction of Inventive Activity Revisited*, edited by Josh Lerner and Scott Stern, 509–14. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226473062.003.0015.

Davis, Fred D. 1986. *Technology Acceptance Model for Empirically Testing New End-User Information Systems Theory and Results*. Unpublished Doctoral Dissertation.

——. 1989. 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology'. *MIS Quarterly* 13 (3): 319. https://doi.org/10.2307/249008.

Davis, Fred D., Richard P. Bagozzi, and Paul R. Warshaw. 1989. 'User Acceptance of Computer Technology: A Comparison of Two Theoretical Models'. *Management Science* 35 (8): 982–1003. https://doi.org/10.1287/mnsc.35.8.982.

Deci, Edward L, and Richard M Ryan. 2000. 'The "What" and "Why" of Goal Pursuits: Human Needs and the Delf-Determination of Behavior'. *Psychological Inquiry* 11 (4): 227–68. https://doi.org/10.1207/S15327965PLI1104.

Decker, Michael, Miltos Ladikas, Susanne Stephan, and Friederike Wütscher, eds. 2004. *Bridges between Science, Society and Policy. Technology Assessment - Methods and Impacts*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-06171-8.

Delvenne, Pierre. 2017. 'Responsible Research and Innovation as a Travesty of Technology

Assessment?' *Journal of Responsible Innovation* 4 (2): 278–88. https://doi.org/10.1080/23299460.2017.1328653.

Deutsche Telekom. 2018. 'Digital Ethics: Guidelines on AI'. 2018. https://www.telekom.com/resource/blob/544508/ca70d6697d35ba60fbcb29aeef4529e8/dl-181008-digitale-ethik-data.pdf.

Deutscher Bundestag. 2020. *Gesetzentwurf Der Bundesregierung: Entwurf Eines Gesetzes Zur Änderung Des Erneuerbare-Energien-Gesetzes Und Weiterer Energierechtlicher Vorschriften (EEG 2021), Drucksache 19/23482*. Germany. https://www.bmwi.de/Redaktion/DE/Downloads/G/gesetzentwurf-aenderung-erneuerbare-energien-gesetzes-und-weiterer-energierechtlicher-vorschriften.pdf?__blob=publicationFile&v=4.

Devine-Wright, Patrick. 2005. 'Beyond NIMBYism: Towards an Integrated Framework for Understanding Public Perceptions of Wind Energy'. *Wind Energy* 8 (2): 125–39. https://doi.org/10.1002/we.124.

——. 2009. 'Rethinking NIMBYism: The Role of Place Attachment and Place Identity in Explaining Place-Protective Action'. *Journal of Community & Applied Social Psychology* 19 (6): 426–41. https://doi.org/10.1002/casp.1004.

——. 2011. 'Place Attachment and Public Acceptance of Renewable Energy: A Tidal Energy Case Study'. *Journal of Environmental Psychology* 31 (4): 336–43. https://doi.org/10.1016/j.jenvp.2011.07.001.

Devine-Wright, Patrick, Susana Batel, Oystein Aas, Benjamin Sovacool, Michael Carnegie Labelle, and Audun Ruud. 2017. 'A Conceptual Framework for Understanding the Social Acceptance of Energy Infrastructure: Insights from Energy Storage'. *Energy Policy* 107 (August): 27–31. https://doi.org/10.1016/j.enpol.2017.04.020.

Devine-Wright, Patrick, and Yuko Howes. 2010. 'Disruption to Place Attachment and the Protection of Restorative Environments: A Wind Energy Case Study'. *Journal of Environmental Psychology* 30 (3): 271–80. https://doi.org/10.1016/j.jenvp.2010.01.008.

DIALOGIK. 2017. 'Partizipationsmodell: Beteiligung an Der Energiewende'. Stuttgart.

Dignum, Virginia. 2018. 'Ethics in Artificial Intelligence: Introduction to the Special Issue'. *Ethics and Information Technology* 20 (1): 1–3. https://doi.org/10.1007/s10676-018-9450-z.

Dignum, Virginia, Matteo Baldoni, Cristina Baroglio, Maurizio Caon, Raja Chatila, Louise Dennis, Gonzalo Génova, et al. 2018. 'Ethics by Design'. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 60–66. New York, NY, USA: ACM. https://doi.org/10.1145/3278721.3278745.

Dincer, Ibrahim. 2012. 'Green Methods for Hydrogen Production'. *International Journal of Hydrogen Energy* 37 (2): 1954–71. https://doi.org/10.1016/j.ijhydene.2011.03.173.

Dodds, Paul E., Iain Staffell, Adam D. Hawkes, Francis Li, Philipp Grünewald, Will McDowall, and Paul Ekins. 2015. 'Hydrogen and Fuel Cell Technologies for Heating: A Review'. *International Journal of Hydrogen Energy* 40 (5): 2065–83. https://doi.org/10.1016/j.ijhydene.2014.11.059.

Dold, Malte F., and Christian Schubert. 2018. 'Toward A Behavioral Foundation of Normative Economics'. *Review of Behavioral Economics* 5 (3–4): 221–41. https://doi.org/10.1561/105.00000097.

Donaldson, Thomas, and Thomas W. Dunfee. 1994. 'Toward A Unified Conception Of Business Ethics: Integrative Social Contracts Theory'. *Academy of Management Review* 19 (2): 252–84. https://doi.org/10.5465/amr.1994.9410210749.

——. 1995. 'Integrative Social Contracts Theory: A Communitarian Conception OfEconomic Ethic'. *Economics and Philosophy* 11 (1): 85–112. https://doi.org/10.1017/S0266267100003230.

Donaldson, Thomas J., and Thomas W. Dunfee. 1999. *Ties That Bind: A Social Contracts Approach to Business Ethics*. Boston: Harvard University Press. http://hdl.handle.net/10822/918895.

Dosi, Giovanni. 2013. 'Innovation, Evolution, and Economics: Where We Are and Where We Should Go'. In *Innovation Studies: Evolution and Future Studies*, edited by Jan Fagerberg, Ben R. Martin, and Esben Sloth Anderson, 111–33. Oxford, United Kingdom: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199686346.003.0005.

Dosi, Giovanni, Chris Freeman, Richard R. Nelson, G. Silverberg, and Luc Soete, eds. 1988. *Technical Change and Economic Theory*. London: Pinter Publishers.

Douglas, Conor M.W., and Dirk Stemerding. 2013. 'Governing Synthetic Biology for Global Health through Responsible Research and Innovation'. *Systems and Synthetic Biology* 7 (3): 139–50. https://doi.org/10.1007/s11693-013-9119-1.

Dreyer, Marc, Luc Chefneux, Anne Goldberg, Joachim von Heimburg, Norberto Patrignani, Monica Schofield, and Chris Shilling. 2017. 'Responsible Innovation: A Complementary View from Industry with Proposals for Bridging Different Perspectives'. *Sustainability* 9 (10): 1–25. https://doi.org/10.3390/su9101719.

Dunfee, Thomas W. 2006. 'A Critical Perspective of Integrative Social Contracts Theory: Recurring Criticisms and next Generation Research Topics'. *Journal of Business Ethics* 68 (3): 303–28. https://doi.org/10.1007/s10551-006-9016-6.

Dunfee, Thomas W. 1991. 'Business Ethics and Extant Social Contracts'. *Business Ethics Quarterly* 1 (1): 23–51. https://doi.org/10.2307/3857591.

Dunfee, Thomas W, and Thomas J Donaldson. 2015. 'Integrative Social Contracts Theory'. In *Wiley Encyclopedia of Management*, 1–5. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118785317.weom020072.

Durant, John. 1999. 'Participatory Technology Assessment and the Democratic Model of the Public Understanding of Science'. *Science and Public Policy* 26 (5): 313–19. https://doi.org/10.3152/147154399781782329.

Dzobo, Kevin, Sampson Adotey, Nicholas E. Thomford, and Witness Dzobo. 2020. 'Integrating Artificial and Human Intelligence: A Partnership for Responsible Innovation in Biomedical Engineering and Medicine'. *OMICS A Journal of Integrative Biology* 24 (5): 247–63. https://doi.org/10.1089/omi.2019.0038.

Earl, Peter E., and Jason Potts. 2004. 'The Market for Preferences'. *Cambridge Journal of Economics* 28: 619–33. https://doi.org/10.1093/cje/beh020.

Ebbeling, Cara B, Dorota B Pawlak, and David S Ludwig. 2002. 'Childhood Obesity: Public-Health Crisis, Common Sense Cure'. *The Lancet* 360 (9331): 473–82. https://doi.org/10.1016/S0140-6736(02)09678-2.

Edler, Jakob. 2006. 'Nachfrageorientierte Innovationspolitik - Politikbrenchmarking'. Berlin. www.tab.fzb.de.

Edler, Jakob, and Jan Fagerberg. 2017. 'Innovation Policy: What, Why, and How'. *Oxford Review of Economic Policy* 33 (1): 2–23. https://doi.org/10.1093/oxrep/grx001.

Elster, Jon. 1983. *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.

Emmerich, Philip, Anna Gesina Hülemeier, David Jendryczko, Manuel Johann Baumann, Marcel Weil, and Dorothee Baur. 2020. 'Public Acceptance of Emerging Energy Technologies in Context of the German Energy Transition'. *Energy Policy* 142 (April): 111516. https://doi.org/10.1016/j.enpol.2020.111516.

Esterberg, Kristin G. 2002. *Qualitative Methods in Social Research*. Boston, Texas: McGrawn Hill Companies.

Etzkowitz, Henry, and Loet Leydesdorff. 2000. 'The Dynamics of Innovation: From National Systems and "Mode 2" to a Triple Helix of University–Industry–Government Relations'. *Research Policy* 29 (2): 109–23. https://doi.org/10.1016/S0048-7333(99)00055-4.

European Commission. 2009. 'Setting up a Community Regime for the Control of Exports, Transfer, Brokering and Transit of Dual-Use Items'. *COUNCIL REGULATION (EC)* 428/2009.

——. 2013. 'Options for Strengthening Responsible Research and Innovation - Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation'. Brussels. https://doi.org/10.2777/46253.

——. 2018. *Mission-Oriented Research Innovation in the European Union: A Problem-Solving Approach to Fuel Innovation-Led Growth*. https://doi.org/10.2777/36546.

——. 2019. *The European Green Deal. COM(2019) 640 Final*. Brussels. https://doi.org/10.1017/CBO9781107415324.004.

——. 2020. 'A Hydrogen Strategy for a Climate-Neutral Europe'. *COM (2020) 301*. Brussels.

Expertenkomission Forschung und Innovation (EFI). 2021. 'Gutachten Zu Forschung, Innovation Und Technologischer Leistungsfähigkeit Deutschlands'. Berlin. http://www.bmbf.de/pubRD/EFI_Gutachten_2015.pdf.

Fagerberg, Jan. 2006. 'Innovation Studies: A Guide to the Literature'. In *The Oxford Handbook of Innovation Management*, edited by Jan Fagerberg, D. Mowery, and Richard R. Nelson. Oxford: Oxford University Press.

Fagerberg, Jan, Ben R. Martin, and Esben Sloth Andersen, eds. 2013. *Innovation Studies: Evolution & Future Studies*. Oxford and New York: Oxford University Press.

Fagerberg, Jan, David C. Mowery, and Richard R. Nelson, eds. 2006. *The Oxford Handbook of Innovation*. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199286805.001.0001.

Fagerberg, Jan, and Bart Verspagen. 2009. 'Innovation Studies—The Emerging Structure of a New Scientific Field'. *Research Policy* 38 (2): 218–33. https://doi.org/10.1016/j.respol.2008.12.006.

Farrell, Alexander E, David W Keith, and James J Corbett. 2003. 'A Strategy for Introducing Hydrogen into Transportation'. *Energy Policy* 31 (13): 1357–67. https://doi.org/10.1016/S0301-4215(02)00195-7.

Fassin, Yves. 2000. 'Innovation and Ethics Ethical: Considerations in the Innovation Business'. *Journal of Business Ethics* 27 (1/2): 193–203. https://doi.org/10.1023/A:1006427106307.

Ferauge, Perrine. 2012. 'A Conceptual Framework of Corporate Social Responsibility and Innovation'. *Global Journal of Business Research* 6 (5): 85–96.

Finkbeiner, Matthias, Erwin M. Schau, Annekatrin Lehmann, and Marzia Traverso. 2010. 'Towards Life Cycle Sustainability Assessment'. *Sustainability* 2 (10): 3309–22. https://doi.org/10.3390/su2103309.

Fisher, E., R. L. Mahajan, and C. Mitcham. 2006. 'Midstream Modulation of Technology:

Governance From Within'. *Bulletin of Science, Technology & Society* 26 (6): 485–96. https://doi.org/10.1177/0270467606295402.

Fisher, Erik. 2020. 'Reinventing Responsible Innovation'. *Journal of Responsible Innovation* 7 (1): 1–5. https://doi.org/10.1080/23299460.2020.1712537.

Flipse, Steven Maarten. 2013. *Enhancing Socially Responsible Innovation in Industry*. TU Delft. https://doi.org/https://doi.org/10.4233/uuid:6d740180-9609-46cb-a403-7904f9a4b1d5.

Floridi, L. 1999. 'Information Ethics: On the Philosophical Foundation of Computer Ethics'. *Ethics and Information Technology* 1: 37–56. https://doi.org/DOI\t10.1023/A:1010018611096.

Floridi, Luciano. 2014. 'Technoscience and Ethics Foresight'. *Philosophy & Technology* 27 (4): 499–501. https://doi.org/10.1007/s13347-014-0180-9.

——. 2018. 'Soft Ethics and the Governance of the Digital'. *Philosophy & Technology*, February, 1–13. https://doi.org/10.1007/s13347-018-0303-9.

——. 2019. 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical'. *Philosophy & Technology* 32 (2): 185–93. https://doi.org/10.1007/s13347-019-00354-x.

Floridi, Luciano, and Josh Cowls. 2019. 'A Unified Framework of Five Principles for AI in Society'. *Harvard Data Science Review* 1 (1): 1–13. https://doi.org/10.1162/99608f92.8cd550d1.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, et al. 2018. 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations'. *Minds and Machines* 28 (4): 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Floridi, Luciano, Josh Cowls, Thomas C. King, and Mariarosaria Taddeo. 2020. 'How to Design AI for Social Good: Seven Essential Factors'. *Science and Engineering Ethics* 26 (3): 1771–96. https://doi.org/10.1007/s11948-020-00213-5.

Flynn, Robert, Miriam Ricci, and Paul Bellaby. 2013. 'Deliberation over New Hydrogen Energy Technologies: Evidence from Two Citizens Panels in the UK'. *Journal of Risk Research* 16 (3–4): 379–91. https://doi.org/10.1080/13669877.2012.743160.

Fraaije, Aafke, and Steven M. Flipse. 2020. 'Synthesizing an Implementation Framework for Responsible Research and Innovation'. *Journal of Responsible Innovation* 7 (1): 113–37. https://doi.org/10.1080/23299460.2019.1676685.

Freeman, R. Edward, Jeffrey S. Harrison, Andrew C. Wicks, Bidhan L. Parmar, and Simone de Colle. 2010. *Stakeholder Theory: The State of the Art*. New York: Cambridge

University Press.

Freeman, R. Edward, Andrew C. Wicks, and Bidhan Parmar. 2004. 'Stakeholder Theory and "The Corporate Objective Revisited"'. *Organization Science* 15 (3): 364–69. https://doi.org/10.1287/orsc.1040.0066.

Freudenburg, William R. 1986. 'Social Impact Assessment'. *Annual Review of Sociology* 12 (1): 451–78. https://doi.org/10.1146/annurev.so.12.080186.002315.

Frey, Carl Benedikt, and Michael A Osborne. 2017. 'The Future of Employment: How Susceptible Are Jobs to Computerisation?' *Technological Forecasting and Social Change* 114 (January): 254–80. http://linkinghub.elsevier.com/retrieve/pii/S0040162516302244.

Friedman, Batya, and David G. Henry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, Massachusetts: MIT Press.

Funtowicz, Silvio O., and Jerome R. Ravetz. 1993. 'The Emergence of Post-Normal Science'. In *Science, Politics and Morality*, edited by René von Schomberg, 85–123. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-015-8143-1_6.

Future of Life Institute. 2018. 'Asilomar AI Principles'. https://futureoflife.org/ai-principles/.

Gallego-Álvarez, Isabel, José Manuel Prado-Lorenzo, and Isabel María García-Sánchez. 2011. 'Corporate Social Responsibility and Innovation: A Resource-Based Theory'. *Management Decision* 49 (10): 1709–27. https://doi.org/10.1108/00251741111183843.

Garden, Hermann, Diana M. Bowman, Sebastian Haesler, and David E. Winickoff. 2016. 'Neurotechnology and Society: Strengthening Responsible Innovation in Brain Science'. *Neuron* 92 (3): 642–46. https://doi.org/10.1016/j.neuron.2016.10.053.

Garden, Hermann, David E. Winickoff, Nina Maria Frahm, and Sebastian Pfotenhauer. 2019. 'Responsible Innovation in Neurotechnology Enterprises'. *OECD Science, Technology and Industry Working Papers* 05: OECD Publishing, Paris. https://doi.org/https://doi.org/10.1787/9685e4fd-en.

Garst, Jilde, Vincent Blok, Léon Jansen, and Onno S.W.F. Omta. 2017. 'Responsibility versus Profit: The Motives of Food Firms for Healthy Product Innovation'. *Sustainability (Switzerland)* 9 (12). https://doi.org/10.3390/su9122286.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, and D B Jan. 2020. 'Datasheets for Datasets', 1–28.

Gehan, E. A. 1965. 'A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples'. *Biometrika* 52 (1–2): 203–24. https://doi.org/10.1093/biomet/52.1-2.203.

Genus, Audley. 2006. 'Rethinking Constructive Technology Assessment as Democratic, Reflective, Discourse'. *Technological Forecasting and Social Change* 73 (1): 13–26.

https://doi.org/10.1016/j.techfore.2005.06.009.

Genus, Audley, and Anne-marie Coles. 2005. 'On Constructive Technology Assessment and Limitations on Public Participation in Technology Assessment'. *Technology Analysis & Strategic Management* 17 (4): 433–43. https://doi.org/10.1080/09537320500357251.

Georgescu-Roegen, Nicholas. 1976. *Energy and Economic Myths*. New York: Pergamon Press. https://doi.org/10.1016/C2013-0-05751-9.

Gibbons, Michael. 2000. 'Mode 2 Society and the Emergence of Context-Sensitive Science'. *Science and Public Policy* 27 (3): 159–63. https://doi.org/10.3152/147154300781782011.

Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott, and Martin Trow. 1994. *The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies*. London: SAGE Publications Ltd.

Gilpin, Leilani H., David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. 2019. 'Explaining Explanations: An Overview of Interpretability of Machine Learning'. *Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA 2018*, 80–89. https://doi.org/10.1109/DSAA.2018.00018.

Giuliani, Elisa. 2018. 'Regulating Global Capitalism amid Rampant Corporate Wrongdoing—Reply to "Three Frames for Innovation Policy"'. *Research Policy* 47 (9): 1577–82. https://doi.org/10.1016/j.respol.2018.08.013.

Godin, Benoit. 2015. 'Innovation: A Study in the Rehabilitation of a Concept'. *Contributions to the History of Concepts* 10 (1): 45–68. https://doi.org/10.3167/choc.2015.100103.

Godin, Benoît. 2008. 'Innovation: The History of a Category'. *Working Paper*.

Godin, Benoît, and Dominique Vinck. 2017. *Critical Studies of Innovation*. Edited by Benoît Godin and Dominique Vinck. Cheltenham, UK: Edward Elgar Publishing. https://doi.org/10.4337/9781785367229.

Goldsmith, Judy, and Emanuelle Burton. 2017. 'Why Teaching Ethics to AI Practitioners Is Important'. *The AAAI-17 Workshop on AI, Ethics, and Society*, 110–14.

Gomes, Leonardo Augusto de Vasconcelos, Ana Lucia Figueiredo Facin, Mario Sergio Salerno, and Rodrigo Kazuo Ikenami. 2018. 'Unpacking the Innovation Ecosystem Construct: Evolution, Gaps and Trends'. *Technological Forecasting and Social Change* 136: 30–48. https://doi.org/10.1016/j.techfore.2016.11.009.

Google. 2018. 'Artificial Intelligence at Google: Our Principles'. https://ai.google/principles/.

Gray, Colin M., Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. 'The Dark (Patterns) Side of UX Design'. *Conference on Human Factors in Computing*

Systems - Proceedings 2018-April: 1–14. https://doi.org/10.1145/3173574.3174108.

Green, Ben. 2019. *The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future*. Cambridge, Massachusetts: MIT Press.

——. 2020. 'Data Science as Political Action: Grounding Data Science in a Politics of Justice'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3658431.

Greene, Daniel, Anna Lauren Hoffmann, and Luke Stark. 2019. 'Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning'. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2122–31. https://doi.org/10.24251/HICSS.2019.258.

Gregory, Terry, Anna Salomons, and Ulrich Zierahn. 2019. 'Racing With or Against the Machine ? Evidence from Europe'.

Grenmmen, Bart, Vincent Blok, and Bernice Bovenkerk. 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): 673–79. https://doi.org/10.1007/s10806-019-09808-w.

Grießhammer, Rainer, Catherine Benoît, Louise Camilla Dreyer, and Anna Flysjö. 2006. 'Feasibility Study: Integration of Social Aspects into LCA'. *UNEP/SETAC Life Cycle Initiative*.

Grießhammer, Rainer, Matthias Buchert, Carl-Otto Gensch, Christian Hochfeld, Andreas Manhart, and Ina Rüdenauer. 2007. 'PROSA–Product Sustainability Assessment (Vol. 49)'. *Öko-Institut e.V.* 49 (0): 222. http://prosa.org/fileadmin/user_upload/pdf/PROSA-gesamt_Finalversion_0407_red.pdf%0Ahttp://www.prosa.org/.

Grinbaum, Alexei, and Christopher Groves. 2013. 'What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues'. In *Responsible Innovation*, edited by Richard Owen, John Bessant, and Maggy Heintz, 119–42. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118551424.ch7.

Gripenberg, Pernilla, Karl-Erik Sveiby, and Beata Segercrantz. 2012. 'Challenging the Innovation Paradigm: The Prevailing Pro-Innovation Bias'. In *Challenging the Innovation Paradigm*, edited by Karl-Erik Sveiby, Pernilla Gripenberg, and Beata Segercrantz. New York: Routledge.

Gross, Catherine. 2007. 'Community Perspectives of Wind Energy in Australia: The Application of a Justice and Community Fairness Framework to Increase Social Acceptance'. *Energy Policy* 35 (5): 2727–36. https://doi.org/10.1016/j.enpol.2006.12.013.

Grossmann, Gene M.;, and Elhanan Helpmann, eds. 1991. *Innovation and Growth in the*

Global Economy. Cambridge, MA: The MIT Press. http://mitpress.mit.edu/catalog/item/default.asp?ttype=2%5C&tid=4444.

Groves, Chris, Lori Frater, Robert Lee, and Elen Stokes. 2011. 'Is There Room at the Bottom for CSR? Corporate Social Responsibility and Nanotechnology in the UK'. *Journal of Business Ethics* 101 (4): 525–52. https://doi.org/10.1007/s10551-010-0731-7.

Grunwald, Armin. 2002. *Technikfolgenabschätzung – Eine Einführung*. Berlin: Ed. Sigma.

——. 2009. 'Technology Assessment: Concepts and Methods'. In *Handbook of the Philosophy of Science*, edited by Anthonie W M Meijers.

——. 2011. 'Responsible Innovation: Bringing Together Technology Assessment, Applied Ethics, and STS Research'. *Enterprise and Work Innovation Studies* 7: 9–31. https://run.unl.pt/bitstream/10362/7944/1/Grunwald9-31.pdf.

——. 2012. 'Responsible Innovation: Neuer Ansatz Der Gestaltung von Technik Und Innovation Oder Nur Ein Schlagwort?' In *Theorie Und Praxis von Technology Governance*, edited by Stephan Bröchler, Georg Aichholzer, and Petra Schaper-Rinkel, 11–22. Wien: (ITA), Institut für Technikfolgen-Abschätzung. http://epub.oeaw.ac.at/ita/ita-manuscript/ita_12_02_Sondernummer.pdf.

——. 2014. 'Technology Assessment for Responsible Innovation'. In *Responsible Innovation 1: Innovative Solutions for Global Issues*, edited by Jeroen Van Den Hoven, Neelke Doorn, Tsjalling Swierstra, Bert Jaap Koops, and Henny Romijn, 15–31. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-8956-1.

Gul, Faruk, and Wolfgang Pesendorfer. 2008. 'The Case for Mindless Economics'. In *The Foundations of Positive and Normative Economics: A Handbook*, edited by A. Caplin and A. Schotter, 3–39. New York: Oxford University Press.

Gunning, David. 2019. 'DARPA's Explainable Artificial Intelligence (XAI) Program'. In *Proceedings of the 24th International Conference on Intelligent User Interfaces - IUI '19*, 20:ii–ii. New York, New York, USA: ACM Press. https://doi.org/10.1145/3301275.3308446.

Gupta, Nidhi, Arnout R.H. Fischer, and Lynn J. Frewer. 2012. 'Socio-Psychological Determinants of Public Acceptance of Technologies: A Review'. *Public Understanding of Science* 21 (7): 782–95. https://doi.org/10.1177/0963662510392485.

Gürses, Seda, Rebekah Overdorf, and Ero Balsa. 2019. 'Stirring The Pots: Protective Optimization Technologies'. In *BEING PROFILED*, edited by Emre Bayamlioglu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt, 24–29. Amsterdam: Amsterdam University Press. https://www.degruyter.com/view/book/9789048550180/10.1515/9789048550180-005.xml.

Guston, D. H. 2014. 'Understanding "Anticipatory Governance"'. *Social Studies of Science* 44 (2): 218–42. https://doi.org/10.1177/0306312713508669.

Guston, David H., Erik Fisher, Armin Grunwald, Richard Owen, Tsjalling Swierstra, and Simone van der Burg. 2014. 'Responsible Innovation: Motivations for a New Journal'. *Journal of Responsible Innovation* 1 (1): 1–8. https://doi.org/10.1080/23299460.2014.885175.

Habermas, Jürgen. 1968. *Technik Und Wissenschaft Als 'Ideologie'*. Frankfurt am Main: Suhrkamp Verlag.

Hagen, H., M. Nitschke, D. Schlindwein, and S. Goll. 2018. 'Acceptance as a Problem, Participation as a Solution? On Premises and Biases in Participatory Research'. In *Technische Unterstützungssysteme, Die Die Menschen Wirklich Wollen*, edited by Robert Weidner and Athanasios Karafillidis, 127–38. Hamburg: Helmut-Schmidt-Universität.

Hagendorff, Thilo. 2020. 'The Ethics of AI Ethics: An Evaluation of Guidelines'. *Minds and Machines* 30 (1): 99–120. https://doi.org/10.1007/s11023-020-09517-8.

Hanekamp, Gerd, and Friederike Wütscher, eds. 2007. *Business Ethics of Innovation*. Ethics of Science and Technology Assessment. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72310-3.

Hansen, Fred, and Michele Smith. 2006. 'The Ethics of Business Strategy'. *Handbook of Business Strategy* 7 (1): 201–6. https://doi.org/10.1108/10775730610618828.

Hansen, Janus. 2006. 'Operationalising the Public in Participatory Technology Assessment: A Framework for Comparison Applied to Three Cases'. *Science and Public Policy* 33 (8): 571–84. https://doi.org/10.3152/147154306781778678.

Harrison, Jeffrey S., and Andrew C. Wicks. 2013. 'Stakeholder Theory, Value, and Firm Performance'. *Business Ethics Quarterly* 23 (1): 97–124. https://doi.org/10.5840/beq20132314.

Hart, Oliver. 2017. 'Incomplete Contracts and Control'. *American Economic Review* 107 (7): 1731–52. https://doi.org/10.1257/aer.107.7.1731.

Hasan, Iftekhar, and Christopher L. Tucci. 2010. 'The Innovation–Economic Growth Nexus: Global Evidence'. *Research Policy* 39 (10): 1264–76. https://doi.org/10.1016/j.respol.2010.07.005.

Hassan, Zaid. 2014. *The Social Labs Revolution: A New Approach to Solving Our Most Complex Challenges*. San Francisco: Berrett-Koehler Publishers.

Haucap, Justus. 2018. *Markt, Macht Und Wettbewerb: Was Steuert Die Datenökonomie*. Berlin: Nicolai Publishing.

Hausman, Daniel M. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.

——. 2012. *Preference, Value, Choice, and Welfare*. Cambridge, United Kingdom: Cambridge University Press.

Hausman, Daniel M., and Michael S. McPherson. 2009. 'Preference Satisfaction and Welfare Economics'. *Economics and Philosophy* 25 (1): 1–25. https://doi.org/10.1017/S0266267108002253.

Hausman, Daniel M., Michael S. McPherson, and Debra Satz. 2017. *Economic Analysis, Moral Philosophy, and Public Policy*. New York: Cambridge University Press.

Häußermann, Johann Jakob. 2019. 'Nudging and Participation: A Contractualist Approach to Behavioural Policy'. *Philosophy of Management*, July. https://doi.org/10.1007/s40926-019-00117-w.

Häußermann, Johann Jakob, and Marie Heidingsfelder. 2017. 'Offen, Verantwortlich Und Verantwortlich Offen'. *TATuP Zeitschrift Für Technikfolgenabschätzung in Theorie Und Praxis* 26 (1–2): 31. https://doi.org/10.14512/tatup.26.1-2.31.

Häußermann, Johann Jakob, Thea Celine Kirsch, Moritz Julian Maier, Simone Kaiser, and Martina Schraudner. 2021. 'Social Acceptance of Green Hydrogen in Germany: Building Trust through Responsible Innovation'. *Energy Policy (Under Review)*.

Häußermann, Johann Jakob, and Christoph Lütge. 2021. 'Community-in-the-Loop: Towards Pluralistic Value Creation in AI, or—Why AI Needs Business Ethics'. *AI and Ethics*, March. https://doi.org/10.1007/s43681-021-00047-2.

Häußermann, Johann Jakob, and Fabian Schroth. 2020. 'Aligning Innovation and Ethics: An Approach to Responsible Innovation Based on Preference Learning'. *Philosophy of Management* 19 (3): 349–64. https://doi.org/10.1007/s40926-019-00120-1.

Heath, Joseph. 2006. 'Business Ethics Without Stakeholders'. Edited by Denis G. Arnold. *Business Ethics Quarterly* 16 (4): 533–57. https://doi.org/10.5840/beq200616448.

——. 2014. *Morality, Competition, and the Firm*. Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199990481.001.0001.

Heath, Joseph, Jeffrey Moriarty, and Wayne Norman. 2010. 'Business Ethics and (or as) Political Philosophy'. Edited by Gary R. Weaver. *Business Ethics Quarterly* 20 (3): 427–52. https://doi.org/10.5840/beq201020329.

Heidingsfelder, Marie, Kora Kimpel, Kathinka Best, and Martina Schraudner. 2015. 'Shaping Future — Adapting Design Know-How to Reorient Innovation towards Public Preferences'. *Technological Forecasting and Social Change* 101 (December): 291–98. https://doi.org/10.1016/j.techfore.2015.03.009.

Heidingsfelder, Marie Lena, Fabian Bitter, and Ronja Ullrich. 2019. 'Debate through Design. Incorporating Contrary Views on New and Emerging Technologies'. *The Design Journal* 22 (sup1): 723–35. https://doi.org/10.1080/14606925.2019.1603658.

Heidingsfelder, Marie Lena, Florian Schütz, and Simone Kaiser. 2016. 'Expanding Participation Participatory Design in Technology Agenda-Setting'. In *Proceedings of the 14th Participatory Design Conference on Short Papers, Interactive Exhibitions, Workshops - PDC '16*, 25–28. New York, New York, USA: ACM Press. https://doi.org/10.1145/2948076.2948087.

Hellman, Deborah. 2020. 'Measuring Algorithmic Fairness'. *Virginia Law Review* 106 (4): 811–66. https://doi.org/arXiv:2001.09784.

Hemphill, Thomas A. 2016. 'Responsible Innovation in Industry: A Cautionary Note on Corporate Social Responsibility'. *Journal of Responsible Innovation* 3 (1): 81–87. https://doi.org/10.1080/23299460.2016.1178896.

Hennen, Leonhard. 1999. 'Participatory Technology Assessment: A Response to Technical Modernity?' *Science and Public Policy* 26 (5): 303–12. https://doi.org/10.3152/147154399781782310.

——. 2012. 'Why Do We Still Need Participatory Technology Assessment?' *Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science* 9 (1–2): 27–41. https://doi.org/10.1007/s10202-012-0122-5.

Hennen, Leonhard, Thomas Petermann, and Constanze Scherz. 2004. 'Partizipative Verfahren Der Technikfolgen-Abschätzung Und Parlamentarische Politikberatung. Neue Formen Der Kommunikation Zwischen Wissenschaft, Politik Und Öffentlichkeit'. *Arbeitsbericht, Büro Für Technikfolgen-Abschätzung Beim Deutschen Bundestag* 96: 1–88.

Heugens, Pursey P. M. A. R., J. van Oosterhout, and Muel Kaptein. 2006. 'Foundations and Applications for Contractualist Business Ethics'. *Journal of Business Ethics* 68 (3): 211–28. https://doi.org/10.1007/s10551-006-9011-y.

Hicks, John R. 1939. 'The Foundations of Welfare Economics'. *The Economic Journal* 49 (196): 696–712. http://www.jstor.org/stable/2234805.

High-Level Expert Group on Artificial Intelligence (HLEG). 2019. 'Ethics Guidelines for Trustworthy AI'. Brussels. https://ec.europa.eu/digital.

Hildebrand, Jan, Caroline Gebauer, and Ansgar Taubitz. 2019. 'Anforderungen an Die Gesellschaftliche Einbettung von Power-to-X Pfaden – Entwicklung Einer Akzeptanzmatrix Als Bewertungsmethodik'. In *Akzeptanz Und Politische Partizipation in Der Energietransformation*, edited by Cornelia Fraune, Michele Knodt, Sebastian Gölzl, and Katharina Langer, 441–61. Wiesbaden: Springer Nature.

https://doi.org/10.1007/978-3-658-24760-7_20.

Hildebrand, Jan, Irina Rau, and Petra Schweizer-Ries. 2012. 'Die Bedeutung Dezentraler Beteiligungsprozesse Für Die Akzeptanz Des Ausbaus Erneuerbarer Energien. Eine Umweltpsychologische Betrachtung'. *Information Zur Raumentwicklung* 9: 491–502.

——. 2018. 'Akzeptanz Und Beteiligung – Ein Ungleiches Paar'. In *Handbuch Energiewende Und Partizipation*, edited by L. Holstenkamp and J. Radtke, 195–209. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-09416-4_12.

Himmelreich, Johannes. 2019. 'Ethics of Technology Needs More Political Philosophy'. *Communications of the ACM* 63 (1): 33–35. https://doi.org/10.1145/3339905.

Hippel, Eric von. 1988. *The Sources of Innovation. International Journal of Forecasting*. Vol. 5. New York & Oxford: Oxford University Press. https://doi.org/10.1016/0169-2070(89)90046-0.

Holbreich, Mark, Jon Genuneit, Juliane Weber, Charlotte Braun-Fahrländer, Marco Waser, and Erika von Mutius. 2012. 'Amish Children Living in Northern Indiana Have a Very Low Prevalence of Allergic Sensitization'. *Journal of Allergy and Clinical Immunology* 129 (6): 1671–73. https://doi.org/10.1016/j.jaci.2012.03.016.

Hollander, Myles, Douglas A. Wolfe, and Eric Chicken. 2015. *Nonparametric Statistical Methods*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119196037.

Holstein, Kenneth, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. 2019. 'Improving Fairness in Machine Learning Systems'. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–16. New York, New York, USA: ACM Press. https://doi.org/10.1145/3290605.3300830.

Holstenkamp, Lars, and Jörg Radtke, eds. 2018. *Handbuch Energiewende Und Partizipation*. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-09416-4.

Hosseini, Seyed Ehsan, and Mazlan Abdul Wahid. 2016. 'Hydrogen Production from Renewable and Sustainable Energy Resources: Promising Green Energy Carrier for Clean Development'. *Renewable and Sustainable Energy Reviews* 57: 850–66. https://doi.org/10.1016/j.rser.2015.12.112.

Hoven, Jeroen van den. 2013. 'Value Sensitive Design and Responsible Innovation'. In *Responsible Innovation*, 75–83. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118551424.ch4.

Hoven, Jeroen van den, Neelke Doorn, Tsjalling Swierstra, Bert-Jaap Koops, and Henny Romijn, eds. 2014. *Responsible Innovation 1*. Dordrecht: Springer Netherlands.

https://doi.org/10.1007/978-94-017-8956-1.

Howaldt, Jürgen, and Michael Schwarz. 2010. *Social Innovation: Concepts, Research Fields, and International Trends. Social Innovation: Concepts, Research Fields, and International Trends*. Dortmund.

Hühn, Matthias P. 2018. 'Responsible Innovation: A Smithian Perspective'. *Philosophy of Management* 17 (1): 41–57. https://doi.org/10.1007/s40926-017-0057-y.

Huijts, N. M.A., E. J.E. Molin, and L. Steg. 2012. 'Psychological Factors Influencing Sustainable Energy Technology Acceptance: A Review-Based Comprehensive Framework'. *Renewable and Sustainable Energy Reviews* 16 (1): 525–31. https://doi.org/10.1016/j.rser.2011.08.018.

Huijts, N. M.A., E. J.E. Molin, and B. van Wee. 2014. 'Hydrogen Fuel Station Acceptance: A Structural Equation Model Based on the Technology Acceptance Framework'. *Journal of Environmental Psychology* 38 (April 2018): 153–66. https://doi.org/10.1016/j.jenvp.2014.01.008.

Husted, Bryan W. 1999. 'A Critique of the Empirical Methods of Integrative Social Contracts Theory'. *Journal of Business Ethics* 20 (3): 227–35. https://doi.org/10.1023/A:1005843230170.

Hyacinth Project. 2017. 'Findings on Public Acceptance'. https://hyacinthproject.eu/wp-content/uploads/2013/12/Findings-on-public-acceptance_short-version__SP_v2.00-1.pdf.

Hytrust. 2013. 'Votum Der Bürgerkonferenz „Mobil Mit Wasserstoff"'. Berlin.

Iatridis, Konstantinos, and Doris Schroeder. 2016. *Responsible Research and Innovation in Industry. Responsible Research and Innovation in Industry: The Case for Corporate Responsibility Tools*. SpringerBriefs in Research and Innovation Governance. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-21693-5.

IEEE. 2017. 'Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems, Version 2'.

Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden. 2016a. '"On the Econ within": A Reply to Daniel Hausman'. *Journal of Economic Methodology* 23 (1): 33–37. https://doi.org/10.1080/1350178X.2015.1070526.

——. 2016b. 'Preference Purification and the Inner Rational Agent: A Critique of the Conventional Wisdom of Behavioural Welfare Economics'. *Journal of Economic Methodology* 23 (1): 1–25. https://doi.org/10.1080/1350178X.2015.1070527.

International Standard Organisation. 2006. *ISO 14040:2006/AMD 1:2020*.

Iofrida, Nathalie, Anna Irene De Luca, Alfio Strano, and Giovanni Gulisano. 2018. 'Can Social Research Paradigms Justify the Diversity of Approaches to Social Life Cycle Assessment?' *International Journal of Life Cycle Assessment* 23 (3): 464–80. https://doi.org/10.1007/s11367-016-1206-6.

Iribarren, Diego, Mario Martín-Gamboa, Javier Manzano, and Javier Dufour. 2016. 'Assessing the Social Acceptance of Hydrogen for Transportation in Spain: An Unintentional Focus on Target Population for a Potential Hydrogen Economy'. *International Journal of Hydrogen Energy* 41 (10): 5203–8. https://doi.org/10.1016/j.ijhydene.2016.01.139.

Ito, Joi. 2018. 'What the Boston School Bus Schedule Can Teach Us About AI An MIT Team Built an Algorithm to Optimize Bell Times and Bus Routes. The Furor around the Plan Offers Lessons in How We Talk to People When We Talk to Them about Artificial Intelligence.' Wired. 2018. https://www.wired.com/story/joi-ito-ai-and-bus-routes/.

Jasanoff, Sheila. 2003. 'Technologies of Humility: Citizen Participation in Governing Science'. *Minerva* 41 (3): 223–44. https://doi.org/10.1023/A:1025557512320.

Jirotka, Marina, Barbara Grimpe, Bernd Stahl, Grace Eden, and Mark Hartswood. 2017. 'Responsible Research and Innovation in the Digital Age'. *Communications of the ACM* 60 (5): 62–68. https://doi.org/10.1145/3064940.

Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. 'The Global Landscape of AI Ethics Guidelines'. *Nature Machine Intelligence* 1 (9): 389–99. https://doi.org/10.1038/s42256-019-0088-2.

Jonas, Hans. 1979. *Das Prinzip Verantwortung*. Frankfurt am Main: Suhrkamp Verlag.

Jong, Irja Marije de, Frank Kupper, Marlous Arentshorst, and Jacqueline Broerse. 2016. 'Responsible Reporting: Neuroimaging News in the Age of Responsible Research and Innovation'. *Science and Engineering Ethics* 22 (4): 1107–30. https://doi.org/10.1007/s11948-015-9684-7.

Joss, Simon. 2002. 'Toward the Public Sphere--Reflections on the Development of Participatory Technology Assessment'. *Bulletin of Science, Technology & Society* 22 (3): 220–31. https://doi.org/10.1177/02767602022003006.

Joss, Simon, and Sergio Bellucci, eds. 2002. *Participatory Technology Assessment: European Perspectives*. London: University of Westminster.

Juma, Calestous. 2016. *Innovation and Its Enemies: Why People Resist New Technologies*. Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467036.001.0001.

Justman, Moshe, and Morris Teubal. 1986. 'Innovation Policy in an Open Economy: A Normative Framework for Strategic and Tactical Issues'. *Research Policy* 15 (3): 121–

38. https://doi.org/10.1016/0048-7333(86)90007-7.

Kahneman, Daniel. 1996. 'Comment (on Plott)'. In *The Rational Foundations of Economic Behaviour*, edited by K. Arrow, E. Colombatto, M. Perlmann, and C. Schmidt, 251–54. Basingstoke.

——. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, Daniel, Paul Slovic, and Amos Tversky, eds. 1982. *Judgement under Uncertainty: Heuristics and Biases*. New York & Cambridge: Cambridge University Press.

Kaiser, Simone, Hannah Glatte, Fabian Bitter, and Marie Heidingsfelder. 2019. 'Foresight as Collaborative Process - Design-Based Scenarios as Strategy Tool in Complex Ecosystems'. In *Vorausschau Und Technologieplanung - 14. Symposium Für Vorausschau Und Technologieplanung*, edited by Jürgen Gausemeier, In Publication. Paderborn: Heinz Nixdorf Institut, Universität Paderborn.

Kalkbrenner, Bernhard J., Koichi Yonezawa, and Jutta Roosen. 2017. 'Consumer Preferences for Electricity Tariffs: Does Proximity Matter?' *Energy Policy* 107 (May): 413–24. https://doi.org/10.1016/j.enpol.2017.04.009.

Kaplan, Andreas, and Michael Haenlein. 2019. 'Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence'. *Business Horizons* 62 (1): 15–25. https://doi.org/10.1016/j.bushor.2018.08.004.

——. 2020. 'Rulers of the World, Unite! The Challenges and Opportunities of Artificial Intelligence'. *Business Horizons* 63 (1): 37–50. https://doi.org/10.1016/j.bushor.2019.09.003.

Kenney, Martin, and John Zysman. 2016. 'The Rise of the Platform Economy'. *Issues in Science and Technology*, 61–69.

Kerr, Anne, Rosemary L. Hill, and Christopher Till. 2018. 'The Limits of Responsible Innovation: Exploring Care, Vulnerability and Precision Medicine'. *Technology in Society* 52: 24–31. https://doi.org/10.1016/j.techsoc.2017.03.004.

Khalil, Omar E.M. 1993. 'Artificial Decision-Making and Artificial Ethics: A Management Concern'. *Journal of Business Ethics* 12 (4): 313–21. https://doi.org/10.1007/BF01666535.

Kieboom, Marlieke. 2014. *Lab Matters: Challenging the Practice of Social Innovation Laboratories*. Amsterdam: Kennisland.

Kleinberg, Jon. 2018. 'Inherent Trade-Offs in Algorithmic Fairness'. In *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*, 40–40. New York, NY, USA: ACM. https://doi.org/10.1145/3219617.3219634.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R. Sunstein. 2018. 'Discrimination in the Age of Algorithms'. *Journal of Legal Analysis* 10 (2005): 1–62. https://doi.org/10.1093/jla/laz001.

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. 2017. 'Inherent Trade-Offs in the Fair Determination of Risk Scores'. *Proceedings of Innovations in Theoretical Computer Science (ITCS)* 67: 1–23. https://doi.org/10.4230/LIPIcs.ITCS.2017.43.

Kloepffer, Walter. 2008. 'Life Cycle Sustainability Assessment of Products'. *International Journal of Life Cycle Assessment* 13 (2): 89–95. https://doi.org/10.1065/lca2008.02.376.

Koirala, Binod Prasad, Ellen van Oost, and Henny van der Windt. 2018. 'Community Energy Storage: A Responsible Innovation towards a Sustainable Energy System?' *Applied Energy* 231 (December): 570–85. https://doi.org/10.1016/j.apenergy.2018.09.163.

Komlos, John. 2016. 'Has Creative Destruction Become More Destructive?' *B.E. Journal of Economic Analysis and Policy* 16 (4). https://doi.org/10.1515/bejeap-2016-0179.

Koops, Bert-Jaap, Ilse Oosterlaken, Henny Romijn, Tsjalling Swierstra, and Jeroen van den Hoven, eds. 2015. *Responsible Innovation 2*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17308-5.

Koops, Bert Jaap, Ilse Oosterlaken, Henny Romijn, Tsjalling Swierstra, and Jeroen van den Hoven. 2015. 'Responsible Innovation 2: Concepts, Approaches, and Applications'. *Responsible Innovation 2: Concepts, Approaches, and Applications*, 1–303. https://doi.org/10.1007/978-3-319-17308-5.

Korinek, Anton, and Joseph Stiglitz. 2017. 'Artificial Intelligence and Its Implications for Income Distribution and Unemployment'. Cambridge, MA. https://doi.org/10.3386/w24174.

Krabbenborg, Lotte, and Henk A.J. Mulder. 2015. *Upstream Public Engagement in Nanotechnology: Constraints and Opportunities*. *Science Communication*. Vol. 37. https://doi.org/10.1177/1075547015588601.

Kulynych, Bogdan, Rebekah Overdorf, Carmela Troncoso, and Seda Gürses. 2018. 'POTs: Protective Optimization Technologies'. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, June, 177–88. https://doi.org/10.1145/3351095.3372853.

Kunin, William E. 2019. 'Robust Evidence of Declines in Insect Abundance and Biodiversity'. *Nature* 574 (7780): 641–42. https://doi.org/10.1038/d41586-019-03241-9.

Lander, Eric S., Françoise Baylis, Feng Zhang, Emmanuelle Charpentier, Paul Berg, Catherine Bourgain, Bärbel Friedrich, et al. 2019. 'Adopt a Moratorium on Heritable Genome Editing'. *Nature* 567 (7747): 165–68. https://doi.org/10.1038/d41586-019-00726-5.

Langer, Katharina, Thomas Decker, and Klaus Menrad. 2017. 'Public Participation in Wind Energy Projects Located in Germany: Which Form of Participation Is the Key to Acceptance?' *Renewable Energy* 112: 63–73. https://doi.org/10.1016/j.renene.2017.05.021.

Langer, Katharina, Thomas Decker, Jutta Roosen, and Klaus Menrad. 2018. 'Factors Influencing Citizens' Acceptance and Non-Acceptance of Wind Energy in Germany'. *Journal of Cleaner Production* 175 (February): 133–44. https://doi.org/10.1016/j.jclepro.2017.11.221.

Larson, Jeff, Surya Mattu, Lauren Kirchner, and Julia Angwin. 2016. 'How We Analyzed the COMPAS Recidivism Algorithm'. ProPublica. 2016. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Larsson, Stefan. 2020. 'On the Governance of Artificial Intelligence through Ethics Guidelines'. *Asian Journal of Law and Society* 00 (October): 1–15. https://doi.org/10.1017/als.2020.19.

Lazar, Seth, and Colin Klein. 2018. 'Why We Need More than Just Data to Create Ethical Driverless Cars'. *The Conversation*, 2018. http://theconversation.com/why-we-need-more-than-just-data-to-create-ethical-driverless-cars-105650.

Leben, Derek. 2017. 'A Rawlsian Algorithm for Autonomous Vehicles'. *Ethics and Information Technology* 19 (2): 107–15. https://doi.org/10.1007/s10676-017-9419-3.

Lecouteux, Guilhem. 2015. *Reconciling Normative and Behavioural Economics. Thèse, École Doctorale de L'École Polytechnique*. Paris.

Lee, Sangil, Erik P. Hess, Christine Lohse, Waqas Gilani, Alanna M. Chamberlain, and Ronna L. Campbell. 2017. 'Trends, Characteristics, and Incidence of Anaphylaxis in 2001-2010: A Population-Based Study'. *Journal of Allergy and Clinical Immunology* 139 (1): 182-188.e2. https://doi.org/10.1016/j.jaci.2016.04.029.

Lehmann, Annekatrin, Eva Zschieschang, Marzia Traverso, Matthias Finkbeiner, and Liselotte Schebek. 2013. 'Social Aspects for Sustainability Assessment of Technologies—Challenges for Social Life Cycle Assessment (SLCA)'. *The International Journal of Life Cycle Assessment* 18 (8): 1581–92. https://doi.org/10.1007/s11367-013-0594-0.

Lenk, Hans, and Matthias Maring. 1993. 'Verantwortung – Normatives Interpretationskonstrukt Und Empirische Beschreibung'. In *Ethische Norm Und Empirische Hypothese*, edited by L.H. Eckensberger and U. Gähde, 222–43. Frankfurt am Main: Suhrkamp Verlag.

Lenk, Hans, and Günter Ropohl, eds. 1987. *Technik Und Ethik*. Stuttgart: Reclam.

Leveringhaus, Alex. 2016. *Ethics and Autonomous Weapons*. London: Palgrave Macmillan

UK. https://doi.org/10.1057/978-1-137-52361-7.

Levine, Hagai, Niels Jørgensen, Anderson Martino-Andrade, Jaime Mendiola, Dan Weksler-Derri, Irina Mindlis, Rachel Pinotti, and Shanna H Swan. 2017. 'Temporal Trends in Sperm Count: A Systematic Review and Meta-Regression Analysis'. *Human Reproduction Update* 23 (6): 646–59. https://doi.org/10.1093/humupd/dmx022.

Liebe, Ulf, Anna Bartczak, and Jürgen Meyerhoff. 2017. 'A Turbine Is Not Only a Turbine: The Role of Social Context and Fairness Characteristics for the Local Acceptance of Wind Power'. *Energy Policy* 107 (May): 300–308. https://doi.org/10.1016/j.enpol.2017.04.043.

Liebert, Wolfgang, and Jan C. Schmidt. 2015. 'Fundierung von „Responsible Research and Innovation" (RRI) Und Normative Ansprüche Prospektiver Technikfolgenabschätzung'. In *Responsible Innovation: Neue Impulse Für Die Technikfolgenabschätzung?*, edited by Alexander Bogner, Michael Decker, and Mahshid Sotoudeh, 1st ed., 379–90. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG. https://doi.org/10.5771/9783845272825-379.

Lienhoop, Nele. 2018. 'Acceptance of Wind Energy and the Role of Financial and Procedural Participation: An Investigation with Focus Groups and Choice Experiments'. *Energy Policy* 118 (March): 97–105. https://doi.org/10.1016/j.enpol.2018.03.063.

Lim, XiaoZhi. 2019. 'Tainted Water: The Scientists Tracing Thousands of Fluorinated Chemicals in Our Environment'. *Nature* 566 (7742): 26–29. https://doi.org/10.1038/d41586-019-00441-1.

Lins, Karl V., Henri Servaes, and Ane Tamayo. 2017. 'Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis'. *Journal of Finance* 72 (4): 1785–1824. https://doi.org/10.1111/jofi.12505.

Lissack, Michael R., and Kurt A. Richardson. 2003. 'Models without Morals: Toward the Ethical Use of Business Models'. *Emergence* 5 (2): 72–102. https://doi.org/10.1207/S15327000EM050208.

Liu, X F, and F N Xu. 2014. 'Corporate Social Responsibility and Innovation: Evidence from Emerging Economies'. In *2014 IEEE International Conference on Management of Innovation and Technology*, 369–74. IEEE. https://doi.org/10.1109/ICMIT.2014.6942455.

Local Energy Consulting. 2020. 'Akzeptanz Und Lokale Teilhabe in Der Energiewende. Handlungsempfehlungen Für Eine Umfassende Akzeptanzpolitik. Impuls Im Auftrag von Agora Energiewende.' Berlin. https://www.agora-energiewende.de/fileadmin2/Projekte/2020/2020_07_EE-Akzeptanz/182_A-EW_Akzeptanz-Energiewende_WEB.pdf.

Loebbecke, Claudia, and Arnold Picot. 2015. 'Reflections on Societal and Business Model Transformation Arising from Digitization and Big Data Analytics: A Research Agenda'. *The Journal of Strategic Information Systems* 24 (3): 149–57. https://doi.org/10.1016/j.jsis.2015.08.002.

Lohmann, Rainer, Knut Breivik, Jordi Dachs, and Derek Muir. 2007. 'Global Fate of POPs: Current and Future Research Directions'. *Environmental Pollution* 150 (1): 150–65. https://doi.org/10.1016/j.envpol.2007.06.051.

Loi, Michele, Andrea Ferrario, and Eleonora Viganò. 2020. 'Transparency as Design Publicity: Explaining and Justifying Inscrutable Algorithms'. *Ethics and Information Technology*, October. https://doi.org/10.1007/s10676-020-09564-w.

Long, Thomas B., and Vincent Blok. 2017. 'Integrating the Management of Socio-Ethical Factors into Industry Innovation: Towards a Concept of Open Innovation 2.0'. *International Food and Agribusiness Management Review* 19 (4): 1–24. https://doi.org/10.22434/IFAMR2017.0040.

Long, Thomas B., Vincent Blok, Steven Dorrestijn, and Phil Macnaghten. 2020. 'The Design and Testing of a Tool for Developing Responsible Innovation in Start-up Enterprises'. *Journal of Responsible Innovation* 7 (1): 45–75. https://doi.org/10.1080/23299460.2019.1608785.

Long, Thomas B., Edurne Iñigo, and Vincent Blok. 2020. 'Responsible Management of Innovation in Business'. In *Research Handbook of Responsible Management*, 606–23. Edward Elgar Publishing. https://doi.org/10.4337/9781788971966.00051.

Louch, Céline, Samuel O. Idowu, and Walter Leal Fllho, eds. 2010. *Innovative CSR: From Risk Management to Value Creation*. London & New York: Routledge. https://doi.org/10.1108/meq.2010.08321eae.001.

Lubberink, Rob. 2018. 'Responsible Innovation in Industry'. University. https://doi.org/10.18174/429296.

Lubberink, Rob, Vincent Blok, Johan van Ophem, and Onno Omta. 2017. 'Lessons for Responsible Innovation in the Business Context: A Systematic Literature Review of Responsible, Social and Sustainable Innovation Practices'. *Sustainability* 9 (5): 721. https://doi.org/10.3390/su9050721.

———. 2019. 'Responsible Innovation by Social Entrepreneurs: An Exploratory Study of Values Integration in Innovations'. *Journal of Responsible Innovation* 6 (2): 179–210. https://doi.org/10.1080/23299460.2019.1572374.

Luca, Anna Irene De, Nathalie Iofrida, Alfio Strano, Giacomo Falcone, and Giovanni Gulisano. 2015. 'Social Life Cycle Assessment and Participatory Approaches: A Methodological Proposal Applied to Citrus Farming in Southern Italy'. *Integrated*

Environmental Assessment and Management 11 (3): 383–96. https://doi.org/10.1002/ieam.1611.

Ludwig, David, and Phil Macnaghten. 2020. 'Traditional Ecological Knowledge in Innovation Governance: A Framework for Responsible and Just Innovation'. *Journal of Responsible Innovation* 7 (1): 26–44. https://doi.org/10.1080/23299460.2019.1676686.

Luetge, Christoph. 2005. 'Economic Ethics, Business Ethics and the Idea of Mutual Advantages'. *Business Ethics: A European Review* 14 (2): 108–18. https://doi.org/10.1111/j.1467-8608.2005.00395.x.

——. 2013. 'The Idea of a Contractarian Business Ethics'. In *Handbook of the Philosophical Foundations of Business Ethics*, 647–58. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-1494-6_24.

——. 2016. 'Contractarian Foundations of Order Ethics'. In *Order Ethics: An Ethical Framework for the Social Market Economy*, 3–17. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-33151-5_1.

Luetge, Christoph, Thomas Armbrüster, and Julian Müller. 2016. 'Order Ethics: Bridging the Gap Between Contractarianism and Business Ethics'. *Journal of Business Ethics* 136 (4): 687–97. https://doi.org/10.1007/s10551-015-2977-6.

Lundvall, Bengt-Åke, ed. 1992. *National Systems of Innovation. Towards a Theory of Innovation and Interactive Learning*. London: Pinter Publishers. https://doi.org/10.26530/oapen_626406.

Lütge, Christoph. 2019. 'There Is Not Enough Business Ethics in the Ethics of Digitization'. In *Ethical Business Leadership in Troubling Times*, edited by Joanne B. Ciulla and Tobey K. Scharding, 280–95. Cheltenham, UK: Edward Elgar Publishing. https://doi.org/10.4337/9781789903058.00026.

Luzzati, T., A. Parenti, and T. Rughi. 2018. 'Economic Growth and Cancer Incidence'. *Ecological Economics* 146 (April): 381–96. https://doi.org/10.1016/j.ecolecon.2017.11.031.

Machiavelli, Niccolò. 1513. *Il Principe*.

Macnaghten, P., R. Owen, J. Stilgoe, B. Wynne, A. Azevedo, A. de Campos, J. Chilvers, et al. 2014. 'Responsible Innovation across Borders: Tensions, Paradoxes and Possibilities'. *Journal of Responsible Innovation* 1 (2): 191–99. https://doi.org/10.1080/23299460.2014.922249.

Madaio, Michael A, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. 'Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI'. In *Proceedings of the 2020 CHI Conference on Human Factors*

in Computing Systems, 1–14. New York, NY, USA: ACM. https://doi.org/10.1145/3313831.3376445.

Madumal, Prashan, Tim Miller, Frank Vetere, and Liz Sonenberg. 2018. 'Towards a Grounded Dialog Model for Explainable Artificial Intelligence'. http://arxiv.org/abs/1806.08055.

Majumdar, Satyajit, Samapti Guha, and Nadiya Marakkath. 2015. *Technology and Innovation for Social Change*. *Technology and Innovation for Social Change*. New Delhi: Springer India. https://doi.org/10.1007/978-81-322-2071-8_1.

Malsch, Ineke, Vrishali Subramanian, Elena Semenzin, Danail Hristozov, Antonio Marcomini, Martin Mullins, Karena Hester, Eamonn McAlea, Finbarr Murphy, and Syed A.M. Tofail. 2015. 'Empowering Citizens in International Governance of Nanotechnologies'. *Journal of Nanoparticle Research* 17 (5): 1–19. https://doi.org/10.1007/s11051-015-3019-0.

Marangunić, Nikola, and Andrina Granić. 2015. 'Technology Acceptance Model: A Literature Review from 1986 to 2013'. *Universal Access in the Information Society* 14 (1): 81–95. https://doi.org/10.1007/s10209-014-0348-1.

Maring, Matthias, ed. 2016. *Zur Zukunft Der Bereichsethiken – Herausforderungen Durch Die Ökonomisierung Der Welt*. Schriftenr. Karlsruhe: KIT Scientific Publishing. https://doi.org/10.5445/KSP/1000054060.

Martin, Ben R. 2012. 'The Evolution of Science Policy and Innovation Studies'. *Research Policy* 41 (7): 1219–39. https://doi.org/10.1016/j.respol.2012.03.012.

Martin, Kirsten. 2019. 'Ethical Implications and Accountability of Algorithms'. *Journal of Business Ethics* 160 (4): 835–50. https://doi.org/10.1007/s10551-018-3921-3.

Martin, Kirsten, Katie Shilton, and Jeffery Smith. 2019. 'Business and the Ethical Implications of Technology: Introduction to the Symposium'. *Journal of Business Ethics* 160 (2): 307–17. https://doi.org/10.1007/s10551-019-04213-9.

Martinez-Conesa, Isabel, Pedro Soto-Acosta, and Mercedes Palacios-Manzano. 2017. 'Corporate Social Responsibility and Its Effect on Innovation and Firm Performance: An Empirical Research in SMEs'. *Journal of Cleaner Production* 142: 2374–83. https://doi.org/10.1016/j.jclepro.2016.11.038.

Martinuzzi, André, Vincent Blok, Alexander Brem, Bernd Stahl, and Norma Schönherr. 2018. 'Responsible Research and Innovation in Industry—Challenges, Insights and Perspectives'. *Sustainability* 10 (3): 702. https://doi.org/10.3390/su10030702.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory*. Oxford and New York: Oxford University Press.

Mathe, Syndhia. 2014. 'Integrating Participatory Approaches into Social Life Cycle Assessment: The SLCA Participatory Approach'. *International Journal of Life Cycle Assessment* 19 (8): 1506–14. https://doi.org/10.1007/s11367-014-0758-6.

Mayring, Philipp. 2007. 'Generalisierung in Qualitativer Forschung'. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 8 (3). ttp://www.qualitati-ve-research.net/fqs-texte/3-07/07-3-26-d.htm.

——. 2010. 'Qualitative Inhaltsanalyse'. In *Handbuch Qualitative Forschung in Der Psychologie*, edited by Günter Mey and Katja Mruck, 601–13. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92052-8_42.

Mayring, Philipp, and Thomas Fenzl. 2019. 'Qualitative Inhaltsanalyse'. In *Handbuch Methoden Der Empirischen Sozialforschung*, edited by Nina Baur and Jörg Blasius, 633–48. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-21308-4_42.

Mazzucato, Mariana. 2016. 'From Market Fixing to Market-Creating: A New Framework for Innovation Policy'. *Industry and Innovation* 23 (2): 140–56. https://doi.org/10.1080/13662716.2016.1146124.

——. 2017. 'Mission-Oriented Innovation Policy: Challenges and Opportunities'. London.

——. 2018. *The Value of Everything: Making and Taking in the Global Economy*. London: Allen Lane.

——. 2021. *Mission Economy: A Moonshot Guide to Changing Capitalism*. London: Allen Lane.

Mazzucato, Mariana, and Carlota Perez. 2014. 'Innovation as Growth Policy: The Challenge for Europe'. *SWPS Working Paper Series , Science Policy Research Unit (SPRU)* 13.

Mccarthy, John, Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon. 2006. 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955'. *AI Magazine* 27 (4): 12. https://doi.org/10.1609/aimag.v27i4.1904.

McCusker, K, and S Gunaydin. 2015. 'Research Using Qualitative, Quantitative or Mixed Methods and Choice Based on the Research'. *Perfusion* 30 (7): 537–42. https://doi.org/10.1177/0267659114559116.

McLennan, Stuart, Amelia Fiske, Leo Anthony Celi, Ruth Müller, Jan Harder, Konstantin Ritt, Sami Haddadin, and Alena Buyx. 2020. 'An Embedded Ethics Approach for AI Development'. *Nature Machine Intelligence* 2 (9): 488–90. https://doi.org/10.1038/s42256-020-0214-1.

McQuillin, Ben, and Robert Sugden. 2012. 'Reconciling Normative and Behavioural Economics: The Problems to Be Solved'. *Social Choice and Welfare* 38 (4): 553–67.

https://doi.org/10.1007/s00355-011-0627-1.

Meadows, Donella H., Dennis L. Meadows, Jorgen Randers, and William W. Behrens III. 1972. 'The Limits to Growth. Report to the Club of Rome'. Rome.

Merriam, Sharan B. 2002. *Qualitative Research in Practice: Examples for Discussion and Analysis*. Edited by Sharan B. Merriam. San Francisco: Jossey-Bass.

Metcalf, Jacob, Emanuel Moss, and Danah Boyd. 2020. 'Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics'. *Social Research: An International Quarterly,* 86 (2): 449–76. https://muse.jhu.edu/article/732185.

Metcalfe, J. S. 2001. 'Institutions and Progress'. *Industrial and Corporate Change* 10 (3): 561–86. https://doi.org/10.1093/icc/10.3.561.

Microsoft. 2018. 'Responsible AI'. 2018. https://www.microsoft.com/en-us/ai/responsible-ai.

Midden, Cees J. H., and Nicole M. A. Huijts. 2009. 'The Role of Trust in the Affective Evaluation of Novel Risks: The Case of CO 2 Storage'. *Risk Analysis* 29 (5): 743–51. https://doi.org/10.1111/j.1539-6924.2009.01201.x.

Mierlo, Barbara van, P. J. Beers, and Anne Charlotte Hoes. 2020. 'Inclusion in Responsible Innovation: Revisiting the Desirability of Opening Up'. *Journal of Responsible Innovation* 7 (3): 361–83. https://doi.org/10.1080/23299460.2020.1780409.

Miles, Morgan P., Linda S. Munilla, and Jeffrey G. Covin. 2004. 'Innovation, Ethics, and Entrepreneurship'. *Journal of Business Ethics* 54 (1): 97–101. https://doi.org/10.1023/B:BUSI.0000043501.13922.00.

Miller, Seumas. 2018. *Dual Use Science and Technology, Ethics and Weapons of Mass Destruction*. SpringerBriefs in Ethics. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-92606-3.

Miller, Tim. 2019. 'Explanation in Artificial Intelligence: Insights from the Social Sciences'. *Artificial Intelligence* 267: 1–38. https://doi.org/10.1016/j.artint.2018.07.007.

Mittelstadt, Brent. 2019. 'Principles Alone Cannot Guarantee Ethical AI'. *Nature Machine Intelligence* 1 (11): 501–7. https://doi.org/10.1038/s42256-019-0114-4.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. 'The Ethics of Algorithms: Mapping the Debate'. *Big Data & Society* 3 (2): 205395171667967. https://doi.org/10.1177/2053951716679679.

Mokyr, Joel. 1990. *The Lever of Riches—Technological Creativity and Economic Progress*. New York & Oxford: Oxford University Press.

——. 1992. 'Technological Inertia in Economic History'. *The Journal of Economic History* 52 (2): 325–38. http://www.jstor.org/stable/2123111.

——. 1994. ‘Progress and Inertia in Technological Change’. In *Capitalism in Context: Essays on Economic Development and Cultural Change in Honor of R.M. Hartwell*, edited by J. A. James and M. Thomas, 230–54. Chicago: University of Chicago Press.

——. 2000. ‘Innovation and Its Enemies: The Economic and Political Roots of Technological Inertia’. In *A Not-so-Dismal Science*, edited by Mancur Olson and Satu Kähköhnen, 61–91. Oxford: Oxford University Press. https://doi.org/10.1093/0198294905.003.0003.

——. 2014. ‘Riding the Technology Dragon’. *Milken Institute Review* 2: 87–94.

Mokyr, Joel, Chris Vickers, and Nicolas L. Ziebarth. 2015. ‘The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?’ *Journal of Economic Perspectives* 29 (3): 31–50. https://doi.org/10.1257/jep.29.3.31.

Montijn-Dorgelo, Fiona N. H., and Cees J. H. Midden. 2008. ‘The Role of Negative Associations and Trust in Risk Perception of New Hydrogen Systems’. *Journal of Risk Research* 11 (5): 659–71. https://doi.org/10.1080/13669870801967218.

Moriarty, Jeffrey. 2005. ‘On the Relevance of Political Philosophy to Business Ethics’. *Business Ethics Quarterly* 15 (3): 455–73.

——. 2017. ‘Business Ethics’. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.). https://plato.stanford.edu/archives/fall2017/entries/ethics-business/.

Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020. ‘From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices’. *Science and Engineering Ethics* 26 (4): 2141–68. https://doi.org/10.1007/s11948-019-00165-5.

Moulaert, F., D. MacCallum, and J. Hillier. 2013. ‘Social Innovation: Intuition, Precept, Concept, Theory and Practice’. In *The International Handbook on Social Innovation: Collective Action, Social Learning and Transdisciplinary Research*, edited by F. Moulaert, D. MacCallum, A. Mehmood, and A. Hamdouch, 13–24. Cheltenham: Edward Elgar.

Moulaert, Frank, Abid Mehmood, Diana MacCallum, and Bernhard Leubolt. 2017. *Social Innovation as a Trigger for Transformations: The Role of Research. European Commission, DG for Research and Innovation*. https://doi.org/10.2777/68949.

Mulgan, Geoff. 2016. ‘Good and Bad Innovation: What Kind of Theory and Practice Do We Need to Distinguish Them?’ London. http://www.nesta.org.uk/sites/default/files/good_and_bad_innovation_by_geoff_mulgan.pdf.

Müller-Wieland, Roda, Katharina Hochfeld, Johann Jakob Häußermann, and Malte Jütting.

2019. 'Kompetenzen Für Eine Digitale Zukunft. Qualifizierungsbedarfe in Den Bereichen Handel, Logistik, Versicherungen Und Krankenkassen'. Stuttgart. http://publica.fraunhofer.de/documents/N-572462.html.

Müller, Wilfried. 1996. 'Technology Assessment: Von Der Abschätzung Ungeahnter Nebenfolgen Zur Bewertung Bekannter Risiken'. In *Schnittstellen*, 59–75. Wiesbaden: Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-663-11337-9_6.

Mumford, John, and David Gray. 2010. 'Consumer Engagement in Alternative Energy—Can the Regulators and Suppliers Be Trusted?' *Energy Policy* 38 (6): 2664–71. https://doi.org/10.1016/j.enpol.2009.05.054.

National Academy of Sciences; National Academy of Medicine; National Academy of Engineering, ed. 2017. *Human Genome Editing*. Washington, D.C.: National Academies Press. https://doi.org/10.17226/24623.

Nelson, Richard R. 1977. *The Moon and the Ghetto: An Essay on Policy Analysis*. New York: W. W. Norton & Company.

———. 1993. *National Innovation Systems: A Comparative Analysis*. Oxford: Oxford University Press.

———. 2011. 'The Moon and the Ghetto Revisited'. *Science and Public Policy* 38 (9): 681–90. https://doi.org/10.3152/030234211X13070021633404.

Noorman, Merel, Tsjalling Swierstra, and Dorien Zandbergen. 2017. 'Questioning the Normative Core of RI: The Challenges Posed to Stakeholder Engagement in a Corporate Setting'. In *Responsible Innovation 3*, 231–49. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-64834-7_13.

Norcia, Vincent di. 1994. 'Ethics, Technology Development, and Innovation'. *Business Ethics Quarterly* 4 (3): 235. https://doi.org/10.2307/3857445.

Nowotny, Helga, Peter B. Scott, and Michael T. Gibbons, eds. 2001. *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. London: Polity Press.

Nowotny, Helga, Peter Scott, and Michael Gibbons. 2003. '"Mode 2" Revisited: The New Production of Knowledge Nine'. *Minerva* 41 (3): 179–94. https://doi.org/10.1023/A:1025505528250.

OECD. 2019. 'Recommendation of the Council on Artificial Intelligence'.

Ogden, Joan M. 1999. 'Prospects for Building a Hydrogen Energy Infrastructure'. *Annual Review of Energy and the Environment* 24 (1): 227–79. https://doi.org/10.1146/annurev.energy.24.1.227.

Oliver, D. G., J. M. Serovich, and T. L. Mason. 2005. 'Constraints and Opportunities with

Interview Transcription: Towards Reflection in Qualitative Research'. *Social Forces* 84 (2): 1273–89. https://doi.org/10.1353/sof.2006.0023.

Otto, Alexander, Martin Robinius, Thomas Grube, Sebastian Schiebahn, Aaron Praktiknjo, and Detlef Stolten. 2017. 'Power-to-Steel: Reducing CO2 through the Integration of Renewable Energy and Hydrogen into the German Steel Industry'. *Energies* 10 (4): 451. https://doi.org/10.3390/en10040451.

Overdest, C., A. Bleicher, and M. Gross. 2010. 'The Experimental Turn in Environmental Sociology: Pragmatism and New Forms of Governance.' In *Environmental Sociology*, edited by M. Gross and H. Heinrichs, 279–94. Dordrecht: Springer.

Owen, R., P. Macnaghten, and J. Stilgoe. 2012. 'Responsible Research and Innovation: From Science in Society to Science for Society, with Society'. *Science and Public Policy* 39 (6): 751–60. https://doi.org/10.1093/scipol/scs093.

Owen, Richard, John Bessant, and Maggy Heintz, eds. 2013. *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*. 1st ed. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118551424.

Owen, Richard, Mario Pansera, Phil Macnaghten, and Sally Randles. 2021. 'Organisational Institutionalisation of Responsible Innovation'. *Research Policy* 50 (1): 104132. https://doi.org/10.1016/j.respol.2020.104132.

Owen, Richard, Jack Stilgoe, Phil Macnaghten, Mike Gorman, Erik Fisher, and Dave Guston. 2013a. 'A Framework for Responsible Innovation'. In *Responsible Innovation. Managing the Responsible Emergence of Science and Innovation in Society*, edited by Richard Owen, Maggy Heintz, and John Bessant, 27–50. Chichester, United Kingdom: John Wiley & Sons, Ltd. https://doi.org/10.1016/j.respol.2013.05.008.

——. 2013b. 'A Framework for Responsible Innovation'. In *Responsible Innovation*, edited by Richard Owen, John Bessant, and Maggy Heintz, 27–50. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118551424.ch2.

Ozarslan, Ahmet. 2012. 'Large-Scale Hydrogen Energy Storage in Salt Caverns'. *International Journal of Hydrogen Energy* 37 (19): 14265–77. https://doi.org/10.1016/j.ijhydene.2012.07.111.

Pacifico Silva, Hudson, Pascale Lehoux, Fiona Alice Miller, and Jean Louis Denis. 2018. 'Introducing Responsible Innovation in Health: A Policy-Oriented Framework'. *Health Research Policy and Systems* 16 (1): 90. https://doi.org/10.1186/s12961-018-0362-5.

Páez, Andrés. 2019. 'The Pragmatic Turn in Explainable Artificial Intelligence (XAI)'. *Minds and Machines* 12 (18). https://doi.org/10.1007/s11023-019-09502-w.

Palazzo, Guido, and Andreas Georg Scherer. 2006. 'Corporate Legitimacy as Deliberation: A

Communicative Framework'. *Journal of Business Ethics* 66 (1): 71–88. https://doi.org/10.1007/s10551-006-9044-2.

Papaioannou, Theo. 2011. 'Technological Innovation, Global Justice and Politics of Development'. *Progress in Development Studies* 11 (4): 321–38. https://doi.org/10.1177/146499341001100404.

Partnership on AI. 2016. 'Tenets'. 2016. https://www.partnershiponai.org/tenets/.

Pellé, Sophie. 2017. *Business, Innovation and Responsibility*. London, Great Britain & Hoboken, USA: ISTE Ltd. & John Wiley & Sons, Inc.

Pellé, Sophie, and Bernard Reber. 2015. 'Responsible Innovation in the Light of Moral Responsibility'. *Journal on Chain and Network Science* 15 (2): 107–17. https://doi.org/10.3920/JCNS2014.x017.

Pellizzone, Anna, Agnes Allansdottir, Roberto De Franco, Giovanni Muttoni, and Adele Manzella. 2015. 'Exploring Public Engagement with Geothermal Energy in Southern Italy: A Case Study'. *Energy Policy* 85 (November 2017): 1–11. https://doi.org/10.1016/j.enpol.2015.05.002.

——. 2017. 'Geothermal Energy and the Public: A Case Study on Deliberative Citizens' Engagement in Central Italy'. *Energy Policy* 101 (September 2016): 561–70. https://doi.org/10.1016/j.enpol.2016.11.013.

Pellizzone, Anna, Agnes Allansdottir, and Adele Manzella. 2019. 'Geothermal Resources in Italy: Tracing a Path Towards Public Engagement'. In , 159–78. https://doi.org/10.1007/978-3-319-78286-7_11.

Pellizzoni, Luigi. 2004. 'Responsibility and Environmental Governance'. *Environmental Politics* 13 (3): 541–65. https://doi.org/10.1080/0964401042000229034.

Perlaviciute, Goda, and Linda Steg. 2014. 'Contextual and Psychological Factors Shaping Evaluations and Acceptability of Energy Alternatives: Integrated Review and Research Agenda'. *Renewable and Sustainable Energy Reviews* 35 (July): 361–81. https://doi.org/10.1016/j.rser.2014.04.003.

Perrault, Raymond, Yoav Shoham, Erik Brynjolfsson, Jack Clark, John Etchemendy, Barbara Grosz Harvard, Terah Lyons, James Manyika, Juan Carlos Niebles, and Saurabh Mishra. 2019. 'The AI Index 2019 Annual Report'. *Standford University - Human-Centered Artificial Intelligence*. Stanford, CA. https://hai.stanford.edu/sites/g/files/sbiybj10986/f/ai_index_2019_report.pdf.

Pesch, Udo. 2012. *Sustainable Innovation, Learning and Responsiblity. Responsible Innovation 1: Innovative Solutions for Global Issues*. https://doi.org/10.1002/9781118551424.

Pessach, Dana, and Erez Shmueli. 2020. 'Algorithmic Fairness'.

Philbeck, Thomas, Nicholas Davis, and Anne Marie Engtoft Larsen. 2018. 'Values, Ethics, and Innovation. Rethinking Technological Development in the Fourth Industrial Revolution.' Geneva.

Phillips, Robert A. 1997. 'Stakeholder Theory and A Principle of Fairness'. *Business Ethics Quarterly* 7 (1): 51–66. https://doi.org/10.2307/3857232.

Phillips, Robert A., and Michael E. Johnson-Cramer. 2006. 'Ties That Unwind: Dynamism in Integrative Social Contracts Theory'. *Journal of Business Ethics* 68 (3): 283–302. https://doi.org/10.1007/s10551-006-9015-7.

Pigou, Arthur C. 1932. *The Economics of Welfare*. London: Macmillan and Co., Ltd.

Poel, Ibo Van de. 2001. 'Investigating Ethical Issues in Engineering Design'. *Science and Engineering Ethics* 7 (3): 429–46. https://doi.org/10.1007/s11948-001-0064-0.

Poel, Ibo van de, Lotte Asveld, Steven Flipse, Pim Klaassen, Zenlin Kwee, Maria Maia, Elvio Mantovani, Christopher Nathan, Andrea Porcari, and Emad Yaghmaei. 2020. 'Learning to Do Responsible Innovation in Industry: Six Lessons'. *Journal of Responsible Innovation* 7 (3): 697–707. https://doi.org/10.1080/23299460.2020.1791506.

Poel, Ibo van de, Lotte Asveld, Steven Flipse, Pim Klaassen, Victor Scholten, and Emad Yaghmaei. 2017. 'Company Strategies for Responsible Research and Innovation (RRI): A Conceptual Model'. *Sustainability* 9 (11): 2045. https://doi.org/10.3390/su9112045.

Preuss, Lutz. 2011. 'Innovative CSR. A Framework for Anchoring Corporate Social Responsibility in the Innovation Literature'. *Journal of Corporate Citizenship* 42 (Summer): 17–33.

Przeworski, Adam. 2019. *Crises of Democracy*. Cambridge & New York: Cambridge University Press. https://doi.org/10.1017/9781108671019.

Pustovit, Svitlana V., and Erin D. Williams. 2010. 'Philosophical Aspects of Dual Use Technologies'. *Science and Engineering Ethics* 16 (1): 17–31. https://doi.org/10.1007/s11948-008-9086-1.

Rahwan, Iyad. 2018. 'Society-in-the-Loop: Programming the Algorithmic Social Contract'. *Ethics and Information Technology* 20 (1): 5–14. https://doi.org/10.1007/s10676-017-9430-8.

Raji, Inioluwa Deborah, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. 'Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing'. https://doi.org/10.1145/3351095.3372873.

Rakova, Bogdana, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2020. 'Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices' 1 (1): 1–18. http://arxiv.org/abs/2006.12358.

Rasmussen, Wiebke. 2018. 'Digitalisierung Braucht Zivilgesellschaft'. Berlin. https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Digitalisierung_braucht_Zivilgesellschaft.pdf.

Rawls, John. 1958. 'Justice as Fairness'. *The Philosophical Review* 67 (2): 164. https://doi.org/10.2307/2182612.

——. 1971. *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press.

——. 2005. *Political Liberalism*. New York: Columbia University Press.

Reisel, William D., and Linda M. Sama. 2003. 'The Distribution of Life-Saving Pharmaceuticals: Viewing the Conflict Between Social Efficiency and Economic Efficiency Through a Social Contract Lens'. *Business and Society Review* 108 (3): 365–87. https://doi.org/10.1111/1467-8594.00169.

Reiss, Julian. 2013. *Philosophy of Economics: A Contemporary Introduction*. New York and London: Routledge.

Renn, Ortwin. 1982. 'Methoden Und Verfahren Der Technikfolgenabschätzung Und Der Technologiebewertung'. In *Technik Auf Dem Prüfstand : Methoden Und Maßstäbe Der Technologiebewertung*, 62–84. Gräfelfing. https://doi.org/10.18419/opus-7233.

Renn, Ortwin, Jürgen Deuschle, Alexander Jäger, and Wolfgang Weimer-Jehle. 2007. *Leitbild Nachhaltigkeit*. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90495-5.

Renssen, Sonja van. 2020. 'The Hydrogen Solution?' *Nature Climate Change* 10 (9): 799–801. https://doi.org/10.1038/s41558-020-0891-0.

Rességuier, Anaïs, and Rowena Rodrigues. 2020. 'AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics'. *Big Data & Society* 7 (2): 1–5. https://doi.org/10.1177/2053951720942541.

Rexhepi, Gadaf, Selma Kurtishi, and Gjilnaipe Bexheti. 2013. 'Corporate Social Responsibility (CSR) and Innovation–The Drivers of Business Growth?' *Procedia - Social and Behavioral Sciences* 75 (April): 532–41. https://doi.org/10.1016/j.sbspro.2013.04.058.

Ribeiro, Barbara E., Robert D. J. Smith, and Kate Millar. 2017. 'A Mobilising Concept? Unpacking Academic Representations of Responsible Research and Innovation'. *Science and Engineering Ethics* 23 (1): 81–103. https://doi.org/10.1007/s11948-016-9761-6.

Ricci, Miriam, Paul Bellaby, and Rob Flynn. 2008. 'What Do We Know about Public Perceptions and Acceptance of Hydrogen? A Critical Review and New Case Study Evidence'. *International Journal of Hydrogen Energy*. https://doi.org/10.1016/j.ijhydene.2008.07.106.

——. 2010. 'Engaging the Public on Paths to Sustainable Energy: Who Has to Trust Whom?' *Energy Policy* 38 (6): 2633–40. https://doi.org/10.1016/j.enpol.2009.05.038.

Rip, Arie. 2014. 'The Past and Future of RRI'. *Life Sciences, Society and Policy* 10 (1): 17. https://doi.org/10.1186/s40504-014-0017-4.

——. 2016. 'The Clothes of the Emperor. An Essay on RRI In and Around Brussels'. *Journal of Responsible Innovation* 0 (0): 1–15. https://doi.org/10.1080/23299460.2016.1255701.

Rip, Arie, Thomas J. Misa, and Johan Schot, eds. 1995. *Managing Technology in Society. The Approach of Constructive Technology Assessment*. London and New York: Pinter Publishers.

Robbins, Lionel. 1932. *An Essay on the Nature and Significane of Economic Science*. London: Macmillan and Co., Ltd.

Rockström, Johan, Will Steffen, Kevin Noone, Åsa Persson, F. Stuart III Chapin, Eric Lambin, Timothy M. Lenton, et al. 2009. 'Planetary Boundaries: Exploring the Safe Operating Space for Humanity'. *Ecology and Society* 14 (2): art32. https://doi.org/10.5751/ES-03180-140232.

Rogers, Everett M. 1962. *Diffusion of Innovations*. New York: The Free Press.

Romer, Paul M. 1986. 'Increasing Returns and Long-Run Growth'. *Journal of Political Economy* 94 (5): 1002–37. https://doi.org/10.1086/261420.

Ropohl, Günter. 1979. *Eine Systemtheorie Der Technik: Zur Grundlegung Der Allgemeinen Technologie*. Munich/Vienna: Hanser.

Rose, David Christian, and Jason Chilvers. 2018. 'Agriculture 4.0: Broadening Responsible Innovation in an Era of Smart Farming'. *Frontiers in Sustainable Food Systems* 2 (December). https://doi.org/10.3389/fsufs.2018.00087.

Rosenbaum, Howard, and Pnina Fichman. 2019. 'Algorithmic Accountability and Digital Justice: A Critical Assessment of Technical and Sociotechnical Approaches'. *Proceedings of the Association for Information Science and Technology* 56 (1): 237–44. https://doi.org/10.1002/pra2.19.

Ruggero, Olga Di. 2014. *Anticipating Public Acceptance: The Hydrogen Case*. Delft: Delft University of Technology. http://www.edgar-program.com/partners/tu-delft/people/55.

Ruggiu, Daniele. 2015. 'Anchoring European Governance: Two Versions of Responsible Research and Innovation and EU Fundamental Rights as "Normative Anchor Points"'. *NanoEthics* 9 (3): 217–35. https://doi.org/10.1007/s11569-015-0240-3.

Ruiz, Cristina, Rosario Marrero, and Bernardo Hernández. 2018. 'Influence of Emotions on the Acceptance of an Oil Drilling Project'. *Environment and Behavior* 50 (3): 324–49. https://doi.org/10.1177/0013916517701795.

Russell, Stuart J., and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach*. Third. Harlow, England: Pearson Education.

Ryan, Mark, and Bernd Carsten Stahl. 2020. 'Artificial Intelligence Ethics Guidelines for Developers and Users: Clarifying Their Content and Normative Implications'. *Journal of Information, Communication and Ethics in Society* ahead-of-p (ahead-of-print). https://doi.org/10.1108/JICES-12-2019-0138.

Saeidi, Sayedeh Parastoo, Saudah Sofian, Parvaneh Saeidi, Sayyedeh Parisa Saeidi, and Seyyed Alireza Saaeidi. 2015. 'How Does Corporate Social Responsibility Contribute to Firm Financial Performance? The Mediating Role of Competitive Advantage, Reputation, and Customer Satisfaction'. *Journal of Business Research* 68 (2): 341–50. https://doi.org/10.1016/j.jbusres.2014.06.024.

Saille, Stevienna de. 2015. 'Innovating Innovation Policy: The Emergence of "Responsible Research and Innovation"'. *Journal of Responsible Innovation* 2 (2): 152–68. https://doi.org/10.1080/23299460.2015.1045280.

Sand, Martin. 2018. 'The Virtues and Vices of Innovators'. In *Futures, Visions, and Responsibility. Technikzukünfte, Wissenschaft Und Gesellschaft / Futures of Technology, Science and Society*, 241–64. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-22684-8_7.

Sanders, Elizabeth B.-N. 2002. 'From Unser-Centered to Participatory Design Approaches'. In *Design and the Social Sciences: Making Connections*, edited by Jorge Frascara, 1–8. New York: Taylor & Francis Ltd.

Sanders, Elizabeth B.-N., Eva Brandt, and Thomas Binder. 2010. 'A Framework for Organizing the Tools and Techniques of Participatory Design'. In *Proceedings of the 11th Biennial Participatory Design Conference on - PDC '10*, 195. New York, New York, USA: ACM Press. https://doi.org/10.1145/1900441.1900476.

Schaer, Philipp. 2017. 'Living Labs –An Ethical Challenge for Researchers and Platform Operators'. In *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*, edited by Michael Zimmer and Katharina Kinder-Kurlanda, 167–76. Peter Lang US. https://doi.org/10.3726/b11077.

Schäfer, Armin, and Michael Zürn. 2021. *Die Demokratische Regression*. Berlin: Suhrkamp

Verlag.

Schäpke, Niko, Matthias Bergmann, Franziska Stelzer, Daniel J. Lang, and Guest Editors. 2018. 'Labs in the Real World: Advancing Transdisciplinary Research and Sustainability Transformation: Mapping the Field and Emerging Lines of Inquiry'. *GAIA - Ecological Perspectives for Science and Society* 27 (1): 8–11. https://doi.org/10.14512/gaia.27.S1.4.

Scharfenberg, David. 2018. 'Computers Can Solve Your Problem. You May Not like the Answer. What Happened When Boston Public Schools Tried for Equity with an Algorithm'. The Boston Globe. 2018. https://apps.bostonglobe.com/ideas/graphics/2018/09/equity-machine/.

Scheidler, Viktoria, and Theresa Pfaff. 2019. 'Wasserstoff in Der Nachbarschaft: Studie Zur Wahrnehmung Und Akzeptanz von H2-Stationen'. *HZwei* 3: 28–30.

Schenk, Anita, Marcel Hunziker, and Felix Kienast. 2007. 'Factors Influencing the Acceptance of Nature Conservation Measures—A Qualitative Study in Switzerland'. *Journal of Environmental Management* 83 (1): 66–79. https://doi.org/10.1016/j.jenvman.2006.01.010.

Scherer, Andreas Georg, and Guido Palazzo. 2007. 'Toward a Political Conception of Corporate Responsibility: Business and Society Seen from a Habermasian Perspective'. *Academy of Management Review* 32 (4): 1096–1120. https://doi.org/10.5465/amr.2007.26585837.

——. 2011. 'The New Political Role of Business in a Globalized World: A Review of a New Perspective on CSR and Its Implications for the Firm, Governance, and Democracy'. *Journal of Management Studies* 48 (4): 899–931. https://doi.org/10.1111/j.1467-6486.2010.00950.x.

Scherhaufer, Patrick, Stefan Höltinger, Boris Salak, Thomas Schauppenlehner, and Johannes Schmidt. 2017. 'Patterns of Acceptance and Non-Acceptance within Energy Landscapes: Acase Study on Wind Energy Expansion in Austria'. *Energy Policy* 109 (October 2016): 863–70. https://doi.org/10.1016/j.enpol.2017.05.057.

Schiederig, Tim, Frank Tietze, and Cornelius Herstatt. 2012. 'Green Innovation in Technology and Innovation Management - an Exploratory Literature Review'. *R&D Management* 42 (2): 180–92. https://doi.org/10.1111/j.1467-9310.2011.00672.x.

Schiff, Daniel, Bogdana Rakova, Aladdin Ayesh, Anat Fanti, and Michael Lennon. 2020. 'Principles to Practices for Responsible AI: Closing the Gap'. http://arxiv.org/abs/2006.04707.

Schillo, R Sandra, and Ryan M Robinson. 2017. 'Inclusive Innovation in Developed Countries: The Who, What, Why, and How'. *Technology Innovation Management Review* 7 (7): 34–46.

https://timreview.ca/sites/default/files/article_PDF/SchilloRobinson_TIMReview_July2017.pdf.

Schlaile, Michael P., Matthias Mueller, Michael Schramm, and Andreas Pyka. 2018. 'Evolutionary Economics, Responsible Innovation and Demand: Making a Case for the Role of Consumers'. *Philosophy of Management* 17 (1): 7–39. https://doi.org/10.1007/s40926-017-0054-1.

Schmidt, Anke, Weert Canzler, and Julia Epp. 2019. 'Welche Rolle Kann Wasserstoff in Der Energie- Und Verkehrswende Spielen?' In *Akzeptanz Und Politische Partizipation in Der Energietransformation. Energietransformation.*, edited by C. Fraune, M. Knodt, S. Gölz, and K. Langer, 419–40. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-24760-7_19.

Schmidt, Jan Cornelius. 2007. 'Normativity and Innovation: An Approach to Concepts of Innovation from the Perspective of Philosophy of Technology'. In *2007 Atlanta Conference on Science, Technology and Innovation Policy*, 1–8. IEEE. https://doi.org/10.1109/ACSTIP.2007.4472880.

Schnellenbach, Jan. 2019. 'Evolving Hierarchical Preferences and Behavioral Economic Policies'. *Public Choice* 178 (1–2): 31–52. https://doi.org/10.1007/s11127-018-0607-4.

Scholl, Gerd, Maike Gossen, Brigitte Holzhauer, and Michael Schipperges. 2015. 'Mit Welchen Kenngrößen Kann Umweltbewusst-Sein Heute Erfasst Werden? – Eine Machbarkeitsstudie'. Berlin.

Schomberg, Rene Von. 2011. 'Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields'. Luxembourg. https://doi.org/10.2139/ssrn.2436399.

Schomberg, René Von. 2012. 'Prospects for Technology Assessment in a Framework of Responsible Research and Innovation'. In *Technikfolgen Abschätzen Lehren – Bildungspotenziale Transdisziplinärer Methoden*, edited by Marc Dusseldorp and Richard Beecroft, 39–61. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-93468-6_2.

——. 2013. 'A Vision of Responsible Research and Innovation'. In *Responsible Innovation. Managing the Responsible Emergence of Science and Innovation in Society*, edited by Richard Owen, John Bessant, and Maggy Heintz, 51–74. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118551424.ch3.

——. 2014. 'The Quest for the "Right" Impacts of Science and Technology: A Framework for Responsible Research and Innovation'. In *Responsible Innovation 1: Innovative Solutions for Global Issues*, edited by Jeroen van den Hoven, Neelke Doorn, Tsjalling Swierstra, Bert-Jaap Koops, and Henny Romijn, 33–50. Dordrecht: Springer

Netherlands. https://doi.org/10.1007/978-94-017-8956-1_3.

Schomberg, René von, and Jonathan Hankins, eds. 2019. *International Handbook on Responsible Innovation: A Global Resource*. Cheltenham: Edward Elgar Publishing. https://doi.org/10.4337/9781784718862.

Schön, Donald A. 1967. *Technology and Change: The Impact of Invention and Innovation on American Social and Economic Development*. New York: Delta Book.

Schormair, Maximilian J. L., and Dirk Ulrich Gilbert. 2020. 'Creating Value by Sharing Values: Managing Stakeholder Value Conflict in the Face of Pluralism through Discursive Justification'. *Business Ethics Quarterly*, June, 1–36. https://doi.org/10.1017/beq.2020.12.

Schot, Johan, and W. Edward Steinmueller. 2018. 'Three Frames for Innovation Policy: R&D, Systems of Innovation and Transformative Change'. *Research Policy* 47 (9): 1554–67. https://doi.org/10.1016/j.respol.2018.08.011.

Schot, Johan W. 1992. 'Constructive Technology Assessment and Technology Dynamics: The Case of Clean Technologies'. *Science, Technology & Human Values* 17 (1): 36–56. https://doi.org/10.1177/016224399201700103.

Schot, Johan W., and Arie Rip. 1997. 'The Past and Future of Constructive Technology Assessment'. *Technological Forecasting and Social Change* 54 (2–3): 251–68. https://doi.org/10.1016/S0040-1625(96)00180-1.

Schroeter, Regina, Oliver Scheel, Ortwin Renn, and Pia-Johanna Schweizer. 2016. 'Testing the Value of Public Participation in Germany: Theory, Operationalization and a Case Study on the Evaluation of Participation'. *Energy Research & Social Science* 13 (March): 116–25. https://doi.org/10.1016/j.erss.2015.12.013.

Schroth, Fabian, Hannah Glatte, and Simone Kaiser. 2020. 'Integrating Civil Society into Regional Innovation Systems - a Social Foresight Lab Approach'. *International Journal of Innovation and Regional Development* 9 (2): 121. https://doi.org/10.1504/IJIRD.2020.110507.

Schroth, Fabian, and Johann Jakob Häußermann. 2018. 'Collaboration Strategies in Innovation Ecosystems: An Empirical Study of the German Microelectronics and Photonics Industries'. *Technology Innovation Management Review* 8 (11): 4–12. https://doi.org/10.22215/timreview/1195.

Schroth, Fabian, and Martina Schraudner. 2019. 'Harnessing and Realizing Social Innovation for RTOs'. In *Atlas of Social Innovation II: Social Innovation and Digital Transformation*, edited by Jürgen Howaldt. Dortmund.

Schubert, Christian. 2009. 'Welfare Creation and Destruction in a Schumpeterian World'.

Papers on Economics and Evolution, No. 0914. http://hdl.handle.net/10419/32638.

——. 2012. 'Is Novelty Always a Good Thing? Towards an Evolutionary Welfare Economics'. *Journal of Evolutionary Economics* 22 (3): 585–619. https://doi.org/10.1007/s00191-011-0257-x.

——. 2013. 'How to Evaluate Creative Destruction: Reconstructing Schumpeter's Approach'. *Cambridge Journal of Economics* 37 (2): 227–50. https://doi.org/10.1093/cje/bes055.

——. 2015a. 'What Do We Mean When We Say That Innovation and Entrepreneurship (Policy) Increase "Welfare"?' *Journal of Economic Issues* 49 (1): 1–22. https://doi.org/10.1080/00213624.2015.1013859.

——. 2015b. 'Opportunity and Preference Learning'. *Economics and Philosophy* 31 (2): 275–95. https://doi.org/10.1017/S0266267115000139.

Schuijff, Mirjam, and Anne M. Dijkstra. 2020. *Practices of Responsible Research and Innovation: A Review. Science and Engineering Ethics*. Vol. 26. Springer Netherlands. https://doi.org/10.1007/s11948-019-00167-3.

Schumpeter, Joseph A. 1934. *The Theory of Economic Development. An Inquiry into Profits, Capital,Credit, Interest, and the Business Cycle*. Cambridge, Massachusetts: Harvard University Press.

——. 1942. *Capitalism, Socialism and Democracy*. New York: Harper & Brothers.

——. 1947. 'The Creative Response in Economic History'. *The Journal of Economic History* 7 (02): 149–59. https://doi.org/10.1017/S0022050700054279.

Schütz, Florian. 2020. 'Das Geschäftsmodell Kollaborativer Innovation : Eine Empirische Analyse Zu Funktionalen Rollen in Quadruple-Helix-Innovationsprozessen'.

Schwab, Klaus. 2016. *The Fourth Industrial Revolution*. New York: Crown Business.

Schweizer-Ries, Petra, Irina Rau, Jan Zoellner, Katrin Nolting, Johannes Rupp, and Dorothee Keppler. 2010. 'Aktivität Und Teilhabe – Akzeptanz Erneuerbarer Energien Durch Beteiligung Steigern'. Magdeburg. https://www.tu-berlin.de/fileadmin/f27/PDFs/Forschung/Abschlussbericht_Aktivitaet_Teilhabe_format.pdf.

Seng, Michelle, and Ah Lee. 2021. 'The Landscape and Gaps in Open Source Fairness Toolkits'.

Sharkey, Noel. 2010. 'Saying "No!" To Lethal Autonomous Targeting'. *Journal of Military Ethics* 9 (4): 369–83. https://doi.org/10.1080/15027570.2010.537903.

Shelley-Egan, Clare, Anders Braarud Hanssen, Laurens Landeweerd, and Bjørn Hofmann. 2018. 'Responsible Research and Innovation in the Context of Human Cognitive

Enhancement: Some Essential Features'. *Journal of Responsible Innovation* 5 (1): 65–85. https://doi.org/10.1080/23299460.2017.1319034.

Shortall, O.K., Sujatha Raman, and Kate Millar. 2015. 'Are Plants the New Oil? Responsible Innovation, Biorefining and Multipurpose Agriculture'. *Energy Policy* 86 (November): 360–68. https://doi.org/10.1016/j.enpol.2015.07.011.

Siebert, Anke, Alberto Bezama, Sinéad O'Keeffe, and Daniela Thrän. 2018. 'Social Life Cycle Assessment: In Pursuit of a Framework for Assessing Wood-Based Products from Bioeconomy Regions in Germany'. *International Journal of Life Cycle Assessment* 23 (3): 651–62. https://doi.org/10.1007/s11367-016-1066-0.

Siebert, Anke, Alberto Bezama, Walther Zeug, Sinead O'Keeffe, Daniela Thrän, and Anne Jäkhel. 2018. 'RESPONSA: REgional SPezifisches CONtextualisiertes Social Life Cycle Assessment'. Leipzig.

Siegrist, Michael, and George Cvetkovich. 2000. 'Perception of Hazards: The Role of Social Trust and Knowledge'. *Risk Analysis* 20 (5): 713–20. https://doi.org/10.1111/0272-4332.205064.

Silva, Luciana Maines da, Claudia Cristina Bitencourt, Kadígia Faccin, and Tatiana Iakovleva. 2019. 'The Role of Stakeholders in the Context of Responsible Innovation: A Meta-Synthesis'. *Sustainability* 11 (6): 1766. https://doi.org/10.3390/su11061766.

Simon, Judith. 2016. 'Value-Sensitive Design and Responsible Research and Innovation'. In *The Ethics of Technology - Methods and Approaches*, edited by Sven Ove Hansson, 219–36. London: Rowman & Littlefield Publishers, Inc.

Singh, Sonal, Shikha Jain, Venkateswaran PS, Avanish K. Tiwari, Mansa R. Nouni, Jitendra K. Pandey, and Sanket Goel. 2015. 'Hydrogen: A Sustainable Fuel for Future of the Transport Sector'. *Renewable and Sustainable Energy Reviews* 51 (November): 623–33. https://doi.org/10.1016/j.rser.2015.06.040.

Soete, Luc. 2013. 'Is Innovation Always Good?' In *Innovation Studies: Evolution and Future Studies*, edited by Jan Fagerberg, Ben R. Martin, and Esben Sloth Andersen, 6:134–44. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199686346.003.0006.

——. 2019. 'Is Innovation Always Good for You? New Policy Challenges for Research and Innovation'. In *International Handbook on Responsible Innovation*, 122–33. Edward Elgar Publishing. https://doi.org/10.4337/9781784718862.00015.

Solow, Robert M. 1957. 'Technical Change and the Aggregate Production Function'. *The Review of Economics and Statistics* 39 (3): 312. https://doi.org/10.2307/1926047.

Soule, Edward. 2002. 'Managerial Moral Strategies—In Search of a Few Good Principles'.

Academy of Management Review 27 (1): 114–24. https://doi.org/10.5465/amr.2002.5922420.

Spiller, Rodger. 2000. 'Ethical Business and Investment: A Model for Business and Society'. *Journal of Business Ethics* 27 (1/2): 149–60. https://doi.org/10.1023/A:1006445915026.

Stahl, Bernd Carsten. 2012. 'Responsible Research and Innovation in Information Systems'. *European Journal of Information Systems* 21 (3): 207–11. https://doi.org/10.1057/ejis.2012.19.

——. 2019. 'Responsible Innovation in ICT: Challenges for Industry'. In *International Handbook on Responsible Innovation*, 367–78. Edward Elgar Publishing. https://doi.org/10.4337/9781784718862.00034.

Stahl, Bernd Carsten, and Mark Coeckelbergh. 2016. 'Ethics of Healthcare Robotics: Towards Responsible Research and Innovation'. *Robotics and Autonomous Systems* 86 (December): 152–61. https://doi.org/10.1016/j.robot.2016.08.018.

Stahl, Bernd Carsten, Grace Eden, and Marina Jirotka. 2013. 'Responsible Research and Innovation in Information and Communication Technology: Identifying and Engaging with the Ethical Implications of ICTs'. In *Responsible Innovation*, 199–218. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118551424.ch11.

Stahl, Bernd Carsten, Job Timmermans, and Catherine Flick. 2016. 'Ethics of Emerging Information and Communication Technologies'. *Science and Public Policy* 44 (September 2016): scw069. https://doi.org/10.1093/scipol/scw069.

Steen, Marc, Martin Sand, and Ibo Van de Poel. 2021. 'Virtue Ethics for Responsible Innovation'. *Business and Professional Ethics Journal* 40 (2): 243–68. https://doi.org/10.5840/bpej2021319108.

Stilgoe, Jack. 2020. *Who's Driving Innovation? Who's Driving Innovation?* https://doi.org/10.1007/978-3-030-32320-2.

Stilgoe, Jack, Richard Owen, and Phil Macnaghten. 2013. 'Developing a Framework for Responsible Innovation'. *Research Policy* 42 (9): 1568–80. https://doi.org/10.1016/j.respol.2013.05.008.

Stirling, Andy. 2006. 'From Science and Society to Science in Society: Towards a Framework for "Co-Operative Research". Report Prepared for European Commission, Directorate-Generale for Research and Innovation'. http://bookshop.europa.eu/en/from-science-and-society-to-science-in-society-pbKINA22072/.

Sugden, Robert. 2015a. 'Looking for a Psychology for the Inner Rational Agent'. *Social Theory and Practice* 41 (4): 579–98. https://doi.org/10.5840/soctheorpract201541432.

——. 2015b. 'Opportunity and Preference Learning: A Reply To Christian Schubert'.

Economics and Philosophy 31 (2): 297–303. https://doi.org/10.1017/S0266267115000140.

Sunstein, Cass R., and Richard H. Thaler. 2003. 'Libertarian Paternalism Is Not an Oxymoron'. *The University of Chicago Law Review* 70 (4): 1159–1202. https://doi.org/10.2307/1600573.

Suškevičs, M., S. Eiter, S. Martinat, D. Stober, E. Vollmer, C.L. de Boer, and M. Buchecker. 2019. 'Regional Variation in Public Acceptance of Wind Energy Development in Europe: What Are the Roles of Planning Procedures and Participation?' *Land Use Policy* 81 (February): 311–23. https://doi.org/10.1016/j.landusepol.2018.10.032.

Sveiby, Karl-Erik, Pernilla Gripenberg, and Beata Segercrantz, eds. 2012. *Challenging the Innovation Paradigm*. Routledge. https://doi.org/10.4324/9780203120972.

Taddeo, Mariarosaria, and Luciano Floridi. 2018. 'How AI Can Be a Force for Good'. *Science* 361 (6404): 751–52. https://doi.org/10.1126/science.aat5991.

Taebi, B., A. Correljé, E. Cuppen, M. Dignum, and U. Pesch. 2014. 'Responsible Innovation as an Endorsement of Public Values: The Need for Interdisciplinary Research'. *Journal of Responsible Innovation* 1 (1): 118–24. https://doi.org/10.1080/23299460.2014.882072.

Takeda, Shutaro, Alexander Keeley, Shigeki Sakurai, Shunsuke Managi, and Catherine Norris. 2019. 'Are Renewables as Friendly to Humans as to the Environment?: A Social Life Cycle Assessment of Renewable Electricity'. *Sustainability* 11 (5): 1370. https://doi.org/10.3390/su11051370.

Tancoigne, Elise, Sally Randles, and Pierre-Benoît Joly. 2016. 'Evolution of a Concept: A Scientometric Analysis of RRI'. In *Navigating Towards Shared Responsibility in Research and Innovation. Approach, Process and Results of the Res-AGorA Project*, edited by Ralf Lindner, 39–44. Karlsruhe: Fraunhofer Institute for Systems and Innovation Research ISI.

Tarde, Gabriel. 1890. *Les Lois de l'imitation*.

——. 1895. *La Logique Sociale*.

——. 1898. *Les Lois Sociales. Esquisse d'une Sociologie*.

——. 1902. *La Psychologie Économique*.

Tasioulas, John. 2019. 'First Steps Towards an Ethics of Robots and Artificial Intelligence'. *Journal of Practical Ethics* 7 (1): 49–83. https://doi.org/10.2139/ssrn.3172840.

Tenner, Edward. 1997. *Why Things Bite Back: Technology and the Revenge of Unintended Consequences*. New York: Vintage Books.

Thaler, Richard H., and Cass R. Sunstein. 2003. 'Libertarian Paternalism'. *The American Economic Review* 93 (2): 175–79.

'The Montreal Declaration'. 2017.

Timmermans, Job, and Vincent Blok. 2018. 'A Critical Hermeneutic Reflection on the Paradigm-Level Assumptions Underlying Responsible Innovation'. *Synthese*, June. https://doi.org/10.1007/s11229-018-1839-z.

Timmermans, Job, Emad Yaghmaei, Bernd Carsten Stahl, and Alexander Brem. 2017. 'Research and Innovation Processes Revisited – Networked Responsibility in Industry'. *Sustainability Accounting, Management and Policy Journal* 8 (3): 307–34. https://doi.org/10.1108/SAMPJ-04-2015-0023.

Tracy, Sarah J. 2010. 'Qualitative Quality: Eight "Big-Tent" Criteria for Excellent Qualitative Research'. *Qualitative Inquiry* 16 (10): 837–51. https://doi.org/10.1177/1077800410383121.

——. 2019. *Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact*. 2nd editio. Hoboken, NJ, USA: John Wiley & Sons, Ltd.

Tsamados, Andreas, Nikita Aggarwal, Josh Cowls, Jessica Morley, Huw Roberts, Mariarosaria Taddeo, and Luciano Floridi. 2020. 'The Ethics of Algorithms: Key Problems and Solutions'. *SSRN Electronic Journal*, 1–32. https://doi.org/10.2139/ssrn.3662302.

Tversky, Amos, and Daniel Kahneman. 1974. 'Judgment under Uncertainty: Heuristics and Biases'. *Science* 185 (4157): 1124–31. https://doi.org/10.1126/science.185.4157.1124.

UK House of Lords, Select Committee on Artificial Intelligence. 2017. 'AI in the UK: Ready, Willing and Able?' London.

United Nations. 1989. 'Montreal Protocol on Substances That Deplete the Ozone Layer (with Annex). Concluded at Montreal on 16 Sep Tember 1987 Authentic Texts : Arabic , Chinese , English , French , Russian and Spanish . MULTILATERAL Protocole de Montréal Relatif' 1522 (26369).

Upham, Paul, Christian Oltra, and Àlex Boso. 2015. 'Towards a Cross-Paradigmatic Framework of the Social Acceptance of Energy Systems'. *Energy Research & Social Science* 8 (July): 100–112. https://doi.org/10.1016/j.erss.2015.05.003.

Valkokari, Katri. 2015. 'Business, Innovation, and Knowledge Ecosystems: How They Differ and How to Survive and Thrive within Them'. *Technology Innovation Management Review* 5 (8): 17–24. https://doi.org/10.22215/timreview/919.

Veale, Michael. 2020. 'A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence'. *European Journal of Risk Regulation*, January, 1–10. https://doi.org/10.1017/err.2019.65.

Veblen, Thorstein. 1898. 'Why Is Economics Not an Evolutionary Science?' *The Quarterly Journal of Economics* 12 (4): 373. https://doi.org/10.2307/1882952.

Venkatesh, Morris, Davis, and Davis. 2003. 'User Acceptance of Information Technology: Toward a Unified View'. *MIS Quarterly* 27 (3): 425. https://doi.org/10.2307/30036540.

Venkatesh, Viswanath, and Fred D. Davis. 2000. 'Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies'. *Management Science* 46 (2): 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926.

Verspagen, Bart. 2006. *Innovation and Economic Growth*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199286805.003.0018.

Verworn, Birgit, and Cornelius Herstatt. 2000. 'Modelle Des Innovationsprozesses'. Working Paper No. 6. Hamburg. http://hdl.handle.net/10419/55484.

Voegtlin, Christian, and Andreas Georg Scherer. 2017. 'Responsible Innovation and the Innovation of Responsibility: Governing Sustainable Development in a Globalized World'. *Journal of Business Ethics* 143 (2): 227–43. https://doi.org/10.1007/s10551-015-2769-z.

Wachter, Sandra, and Brent Mittelstadt. 2019. 'A Right to Resasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI'. *Colum. Bus. L. Rev.* Forthcomin (2): 1–84.

Wachter, Sandra, Brent Mittelstadt, and Chris Russell. 2018. 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR'. *Harvard Journal of Law and Technology* 31 (2). http://arxiv.org/abs/1606.08813.

Wagner, Ben. 2018. 'Ethics as an Escape from Regulation: From "ethics-Washing" to Ethics-Shopping?' In *Being Profiled. Cogitas Ergo Sum*, edited by M. Hildebrandt, 84–90. Amsterdam: Amsterdam University Press.

Ward, Ken. 2018. 'Social Networks, the 2016 US Presidential Election, and Kantian Ethics: Applying the Categorical Imperative to Cambridge Analytica's Behavioral Microtargeting'. *Journal of Media Ethics* 33 (3): 133–48. https://doi.org/10.1080/23736992.2018.1477047.

Wempe, Ben. 2004. 'On the Use of the Social Contract Model in Business Ethics'. *Business Ethics: A European Review* 13 (4): 332–41. https://doi.org/10.1111/j.1467-8608.2004.00374.x.

——. 2008. 'Four Design Criteria for Any Future Contractarian Theory of Business Ethics'. *Journal of Business Ethics* 81 (3): 697–714. https://doi.org/10.1007/s10551-007-9542-x.

——. 2009. 'Extant Social Contracts and the Question of Business Ethics'. *Journal of Business Ethics* 88 (SUPPL. 4): 741–50. https://doi.org/10.1007/s10551-009-0326-3.

Whitman, Douglas Glen, and Mario J. Rizzo. 2015. 'The Problematic Welfare Standards of Behavioral Paternalism'. *Review of Philosophy and Psychology* 6 (3): 409–25. https://doi.org/10.1007/s13164-015-0244-5.

Whittaker, M., Crawford, K., Dobbe, R. et al. 2018. 'AI Now Report 2018'. *AI Now*, no. December. https://ainowinstitute.org/AI_Now_2018_Report.pdf.

Whittlestone, Jess, Anna Alexandrova, Rune Nyrup, and Stephen Cave. 2019. 'The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions'. *AIES 2019 - Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 195–200. https://doi.org/10.1145/3306618.3314289.

Wilkinson, Sue. 1998. 'Focus Group Methodology: A Review'. *International Journal of Social Research Methodology* 1 (3): 181–203. https://doi.org/10.1080/13645579.1998.10846874.

Wilsdon, James, and Rebecca Willis. 2004. *See-Through Science: Why Public Engagement Needs to Move Upstream*. London: Demos. https://doi.org/10.1016/S0278-4254(04)00022-5.

Winner, Langdon. 2018. 'The Cult of Innovation: Its Myths and Rituals'. In *Engineering a Better Future*, 51:61–73. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-91134-2_8.

Witt, Ulrich. 1996. 'Innovations, Externalities and the Problem of Economic Progress'. *Public Choice* 89: 113–30. https://doi.org/10.1007/BF00114282.

——. 2001. 'Learning to Consume - A Theory of Wants and the Growth of Demand'. *Journal of Evolutionary Economics* 11 (1): 23–36. https://doi.org/10.1007/PL00003851.

——. 2016. 'What Kind of Innovations Do We Need to Secure Our Future?' *Journal of Open Innovation: Technology, Market, and Complexity* 2 (3). https://doi.org/10.1186/s40852-016-0043-y.

Wolsink, Maarten. 2000. 'Wind Power and the NIMBY-Myth: Institutional Capacity and the Limited Significance of Public Support'. *Renewable Energy* 21 (1): 49–64. https://doi.org/10.1016/S0960-1481(99)00130-5.

——. 2007. 'Planning of Renewables Schemes: Deliberative and Fair Decision-Making on Landscape Issues Instead of Reproachful Accusations of Non-Cooperation'. *Energy Policy* 35 (5): 2692–2704. https://doi.org/10.1016/j.enpol.2006.12.002.

Wolsink, Maarten, and Sylvia Breukers. 2010. 'Contrasting the Core Beliefs Regarding the Effective Implementation of Wind Power. An International Study of Stakeholder Perspectives'. *Journal of Environmental Planning and Management* 53 (5): 535–58. https://doi.org/10.1080/09640561003633581.

Wong, Pak-hang. 2020. 'Democratizing Algorithmic Fairness'. *Philosophy & Technology* 33 (2): 225–44. https://doi.org/10.1007/s13347-019-00355-w.

Woodhouse, Edward, and Daniel Sarewitz. 2007. 'Science Policies for Reducing Societal Inequities'. *Science and Public Policy* 34 (2): 139–50. https://doi.org/10.3152/030234207X195158.

Wu, Ruqun, Dan Yang, and Jiquan Chen. 2014. 'Social Life Cycle Assessment Revisited'. *Sustainability* 6 (7): 4200–4226. https://doi.org/10.3390/su6074200.

Wu, Tim. 2016. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. New York: Alfred A. Knopf.

Wüstenhagen, Rolf, Maarten Wolsink, and Mary Jean Bürer. 2007. 'Social Acceptance of Renewable Energy Innovation: An Introduction to the Concept'. *Energy Policy* 35 (5): 2683–91. https://doi.org/10.1016/j.enpol.2006.12.001.

Wynsberghe, Aimee van, and Scott Robbins. 2014. 'Ethicist as Designer: A Pragmatic Approach to Ethics in the Lab'. *Science and Engineering Ethics* 20 (4): 947–61. https://doi.org/10.1007/s11948-013-9498-4.

Yaghmaei, Emad, and Ibo Van De Poel. 2020. *Assessment of Responsible Innovation*. Edited by Emad Yaghmaei and Ibo van de Poel. *Assessment of Responsible Innovation*. Routledge. https://doi.org/10.4324/9780429298998.

Yang, Kai Cheng, Onur Varol, Clayton A. Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2019. 'Arming the Public with Artificial Intelligence to Counter Social Bots'. *Human Behavior and Emerging Technologies* 1 (1): 48–61. https://doi.org/10.1002/hbe2.115.

Yeung, Karen. 2017. '"Hypernudge": Big Data as a Mode of Regulation by Design'. *Information, Communication & Society* 20 (1): 118–36. https://doi.org/10.1080/1369118X.2016.1186713.

Yoshinaka, Yutaka, Christian Clausen, and Annegrethe Hansen. 2003. 'The Social Shaping of Technology: A New Space for Politics?' In *Technikgestaltung: Zwischen Wunsch Oder Wirklichkeit*, edited by Armin Grunwald, 117–31. Berlin: Springer.

Zachariah-Wolff, J. Leslie, and Kas Hemmes. 2006. 'Public Acceptance of Hydrogen in the Netherlands: Two Surveys That Demystify Public Views on a Hydrogen Economy'. *Bulletin of Science, Technology & Society* 26 (4): 339–45. https://doi.org/10.1177/0270467606290308.

Zeng, Daniel, Hsinchun Chen, Robert Lusch, and Shu-Hsing Li. 2010. 'Social Media Analytics and Intelligence'. *IEEE Intelligent Systems* 25 (6): 13–16. https://doi.org/10.1109/MIS.2010.151.

Zhang, Fan, Pengcheng Zhao, Meng Niu, and Jon Maddy. 2016. 'The Survey of Key Technologies in Hydrogen Energy Storage'. *International Journal of Hydrogen Energy* 41 (33): 14535–52. https://doi.org/10.1016/j.ijhydene.2016.05.293.

Zhu, Haiyi, Bowen Yu, Aaron Halfaker, and Loren Terveen. 2018. 'Value-Sensitive Algorithm Design: Method, Case Study, and Lessons'. *Proceedings of the ACM on Human-Computer Interaction* 2 (CSCW): 1–23. https://doi.org/10.1145/3274463.

Ziegler, Rafael. 2015. 'Justice and Innovation – Towards Principles for Creating a Fair Space for Innovation'. *Journal of Responsible Innovation* 2 (2): 184–200. https://doi.org/10.1080/23299460.2015.1057796.

Zimmer, René, and Jörg Welke. 2012. 'Let's Go Green with Hydrogen! The General Public's Perspective'. *International Journal of Hydrogen Energy* 37 (22): 17502–8. https://doi.org/10.1016/j.ijhydene.2012.02.126.

Zoellner, Jan, Petra Schweizer-Ries, and Christin Wemheuer. 2008. 'Public Acceptance of Renewable Energies: Results from Case Studies in Germany'. *Energy Policy* 36 (11): 4136–41. https://doi.org/10.1016/j.enpol.2008.06.026.

APPENDIX

Copyright Clearance Center Marketplace™

Springer Nature BV - License Terms and Conditions

This is a License Agreement between Johann Jakob Haeussermann ("You") and Springer Nature BV ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer Nature BV, and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	03-May-2021	Type of Use	Republish in a thesis/dissertation
Order license ID	1116264-1		
ISSN	2052-9597	Publisher	Springer
		Portion	Chapter/article

LICENSED CONTENT

Publication Title	Philosophy of Management	Rightsholder	Springer Nature BV
		Publication Type	e-Journal
Article Title	Aligning Innovation and Ethics: an Approach to Responsible Innovation Based on Preference Learning	Start Page	349
		End Page	364
		Issue	3
		Volume	19
Author/Editor	Libri Publishing		
Date	01/01/2015		
Language	English		
Country	Germany		

REQUEST DETAILS

Portion Type	Chapter/article	Rights Requested	Main product
Page range(s)	340-364	Distribution	Worldwide
Total number of pages	16	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the disabled?	No
Who will republish the content?	Academic institution	Minor editing privileges?	No
		Incidental promotional use?	No
Duration of Use	Life of current edition		
Lifetime Unit Quantity	Up to 499	Currency	EUR

NEW WORK DETAILS

Title	Responsible Innovation, PhD Thesis	Institution name	TU Munich
		Expected presentation date	2021-12-01
Instructor name	Johann Jakob Haeussermann		

ADDITIONAL DETAILS

The requesting person / organization to appear on the license	Johann Jakob Haeussermann

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Aligning Innovation and Ethics: an Approachto Responsible Innovation Basedon Preference Learning	**Title of the article/chapter the portion is from**	Aligning Innovation and Ethics: an Approach to Responsible Innovation Based on Preference Learning
Editor of portion(s)	Häußermann, Johann Jakob; Schroth, Fabian	**Author of portion(s)**	Häußermann, Johann Jakob; Schroth, Jakob
Volume of serial or monograph	19	**Publication date of portion**	2019-09-12
Page or page range of portion	349-364		

PUBLISHER TERMS AND CONDITIONS

If you are placing a request on behalf of/for a corporate organization, please use RightsLink. For further information visit http://www.nature.com/reprints/permission-requests.html and https://www.springer.com/gp/rights-permissions/obtaining-permissions/882. If the content you are requesting to reuse is under a CC-BY 4.0 licence (or previous version), you do not need to seek permission from Springer Nature for this reuse as long as you provide appropriate credit to the original publication. https://creativecommons.org/licenses/by/4.0/

CCC Republication Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

    3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

    3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291.

Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed

the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

    8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

    8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

    8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

    8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

    8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any

Copyright Clearance Center Marketplace™

Springer - License Terms and Conditions

This is a License Agreement between Johann Jakob Haeussermann ("You") and Springer ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	03-May-2021	Type of Use	Republish in a thesis/dissertation
Order license ID	1116264-2		
ISSN	2730-5953	Publisher	Springer Nature
		Portion	Chapter/article

LICENSED CONTENT

Publication Title	AI and Ethics	Rightsholder	Springer
Article Title	Community-in-the-loop: towards pluralistic value creation in AI, or—why AI needs business ethics	Publication Type	Journal
Date	01/01/2020		

REQUEST DETAILS

Portion Type	Chapter/article	Rights Requested	Main product
Page range(s)	1-22	Distribution	Worldwide
Total number of pages	22	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the disabled?	No
Who will republish the content?	Academic institution	Minor editing privileges?	No
Duration of Use	Life of current edition	Incidental promotional use?	No
Lifetime Unit Quantity	Up to 499	Currency	EUR

NEW WORK DETAILS

Title	Responsible Innovation, PhD Thesis	Institution name	TU Munich
		Expected presentation date	2021-12-01
Instructor name	Johann Jakob Haeussermann		

ADDITIONAL DETAILS

Order reference number	N/A	The requesting person / organization to appear on the license	Johann Jakob Haeussermann

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Community-in-the-loop: towards pluralistic value creation in AI, or—why AI needs business ethics	**Title of the article/chapter the portion is from**	Community-in-the-loop: towards pluralistic value creation in AI, or—why AI needs business ethics
Editor of portion(s)	Häußermann, Johann Jakob; Lütge, Christoph	**Author of portion(s)**	Häußermann, Johann Jakob; Lütge, Christoph
Volume of serial or monograph	N/A	**Issue, if republishing an article from a serial**	N/A
Page or page range of portion	1-22	**Publication date of portion**	2021-03-23

PUBLISHER TERMS AND CONDITIONS

A maximum of 10% of the content may be licensed for republication. The user is responsible for identifying and seeking separate licenses for any third party materials that are identified anywhere in the work. Without a separate license, such third party materials may not be reused.
If you are placing a request on behalf of/for a corporate organization, please use RightsLink. For further information visit http://www.nature.com/reprints/permission-requests.html and https://www.springer.com/gp/rights-permissions/obtaining-permissions/882. If the content you are requesting to reuse is under a CC-BY 4.0 licence (or previous version), you do not need to seek permission from Springer Nature for this reuse as long as you provide appropriate credit to the original publication. https://creativecommons.org/licenses/by/4.0/

CCC Republication Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

    3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

    3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of

the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL

OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

    8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

    8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

    8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

    8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

    8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

Häußermann, Johann Jakob

Von: Energy Policy <energypolicy@elsevier.com>
Gesendet: Dienstag, 20. April 2021 07:33
An: Häußermann, Johann Jakob
Betreff: Re: Author query JEPO JEPO-D-21-00241 [210420-002837]

Dear Dr. Jakob Haeussermann,

Thank you for the email. Yes. You could include your article in full or in part in a thesis or dissertation for non-commercial purposes.

Please do not hesitate to contact me if I can be of any further assistance.

Kind Regards,
Vimok

Vimok Bellan
Journal Manager
ELSEVIER | Operations | Research Content Operations
+91 44 4299 4836 office
v.bellanjoghi@elsevier.com
www.elsevier.com

From: Johann Jakob Häußermann
Date: Monday, April 19, 2021 09:27 PM GMT

Please enter your name, e-mail address and query below, then click 'Preview and Send'.

Name: Jakob Haeussermann
E-mail address: johann-jakob.haeussermann@iao.fraunhofer.de

What is the current status of your paper: Under Review_x_ Accept__ Reject__ Revise__

Query:

Dear ladies and gentlemen,

herewith I would like to ask about the possibility to re-use the manuscript of my paper currently under review as chapter of my PhD dissertation. Would it be possible to re-print it in my thesis with the addition as being currently under review at your journal? Or would that hinder the publication process at Energy Policy?

Thanks a lot for your help!

Best regards,

Please do not modify the text below this line.

From Corresponding author: Mr. Johann Jakob Häußermann

Journal title: Energy Policy

Article title: Social acceptance of green hydrogen in Germany: building trust through responsible innovation

Manuscript number: JEPO-D-21-00241

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: https://www.editorialmanager.com/jepo/login.asp?a=r). Please contact the publication office if you have any questions.
