UNIVERSITÄT
ZU KÖLN

Toward a Responsible Design Science Research Ecosystem for the Digital Age

A Critical Pragmatist Perspective

BY ALEXANDER HERWIX

Toward a Responsible Design Science Research Ecosystem for the Digital Age: A Critical Pragmatist Perspective

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2024

presented
by

M.Sc. Alexander Herwix

from

Willich, Germany

First reviewer: Prof. Dr. Christoph Rosenkranz
Second reviewer: Prof. Dr. Detlef Schoder
Date of oral defense: 22.04.2024

Zusammenfassung

Diese Dissertation ist motiviert durch die Notwendigkeit, Wege zu finden, verantwortungsvoll durch die komplexe Landschaft des digitalen Zeitalters zu navigieren, in der die rasanten Fortschritte in der Informationstechnologie (IT), insbesondere im Bereich der künstlichen Intelligenz (KI), sowohl beispiellose Chancen als auch potenziell katastrophale Risiken für die Gesellschaft mit sich bringen. Ausgangspunkt ist die Annahme, dass wir uns auf verantwortungsvolle Innovation (Responsible Innovation, RI) konzentrieren müssen, um die Vorteile der sich beschleunigenden IT-Innovation zu nutzen und gleichzeitig die gefährlichsten Risiken zu vermeiden.

Das Hauptziel dieser Dissertation ist es, die Entwicklung eines Ökosystems für verantwortungsvolle Design Science Research (DSR) zu initiieren und zu unterstützen, um die DSR mit den Imperativen von RI in Einklang zu bringen. Zu diesem Zweck befürwortet sie einen Paradigmenwechsel für die Forschung im Bereich der Informationssysteme (IS) und schlägt verantwortungsvolle DSR als "Supermethodologie" vor, um die großen Herausforderungen des digitalen Zeitalters auf verantwortungsvolle und produktive Weise anzugehen. Die Dissertation erreicht dieses Ziel durch drei miteinander verbundene Studien:

1. Ethische Grundlagen der IS-Forschung: Ein umfassender Überblick über Ethik und ihre Verbindungen zur IS-Forschung, basierend auf einer Podiumsdiskussion und Literaturrecherche, zeigt ethische Überlegungen für eine verantwortungsvolle IS-Forschung im digitalen Zeitalter auf. Dies schafft die Grundlage für eine bewusste Auseinandersetzung mit RI in der IS-Forschung und DSR.

2. Konzeptioneller Rahmen für verantwortungsvolle DSR: Der zentrale Teil der Dissertation entwickelt eine fundierte (multi-grounded) Theorie, die durch Interviews mit verschiedenen Mitgliedern der IS-Community und akademischer Literatur untermauert wird. Das Ergebnis dieser Untersuchung ist ein ganzheitlicher Rahmen für verantwortungsvolle DSR, der grundlegende paradigmatische Herausforderungen aufgreift: ontologische, epistemologische, axiologische und methodologische. Dieser dient als offene und integrative konzeptionelle Grundlage für ein verantwortungsvolles DSR-Ökosystem.

3. Angewandtes verantwortungsvolles DSR-Projekt: Die Dissertation schließt mit einem konkreten Projekt ab, das als Beispiel dafür dient, wie verantwortungsvolle DSR in der Praxis aussehen kann. Mithilfe von Szenarioentwicklung und System Dynamics basierten Simulationsexperimenten werden die Auswirkungen der Digitalisierung auf die Resilienz des Lebensmittelsektors in den USA in katastrophalen Stromausfallszenarien untersucht. Die Studie regt einen kritischen Diskurs über das Zusammenspiel von digitaler Transformation und gesellschaftlicher Resilienz an und liefert Erkenntnisse für eine verantwortungsvolle Einbettung digitaler Technologien.

Insgesamt bilden diese Studien die Grundlage für ein verantwortungsvolles DSR-Ökosystem. Diese Dissertation erkennt jedoch den partizipativen und sich entwickelnden Charakter solcher Entwicklungsbemühungen an und versteht sich daher nur als Ausgangspunkt. Es bedarf eines kontinuierlichen Diskurses und bewussten Handelns, um verantwortungsvolle DSR voranzutreiben und zum Gedeihen unserer Gesellschaften im digitalen Zeitalter beizutragen.

Abstract

This dissertation is motivated by the need to find ways to responsibly navigate the complex landscape of the digital age, in which rapid advances in information technology (IT), particularly in the field of artificial intelligence (AI), present both unprecedented opportunities and potentially catastrophic risks to society. The starting point is the assumption that we need to focus on Responsible Innovation (RI) in order to reap the benefits of accelerating IT innovation while avoiding the most dangerous risks.

The main goal of this dissertation is to initiate and support the development of a responsible design science research (DSR) ecosystem to align DSR with the imperatives of RI. To this end, it advocates a paradigm shift for Information Systems (IS) research and proposes responsible DSR as a "supermethodology" to address the grand challenges of the digital age in a responsible and productive manner. This dissertation achieves this goal through three interrelated inquiries:

1. Ethical Foundations of IS Research: A comprehensive overview of ethics and its relationship to IS research, based on a panel discussion and literature review, highlights ethical considerations for responsible IS research in the digital age. This sets the stage for a conscious engagement with RI in IS research and DSR.

2. Conceptual Framework for Responsible DSR: The central part of the dissertation develops a multi-grounded theory supported by interviews with various members of the IS community and academic literature. The result is a holistic framework for responsible DSR that addresses fundamental paradigmatic challenges: ontological, epistemological, axiological, and methodological. The framework serves as an open and integrative conceptual foundation for a responsible DSR ecosystem.

3. Applied Responsible DSR Project: The dissertation concludes with a concrete responsible DSR project that not only contributes to addressing a relevant societal problem, but also serves as an example of what responsible DSR can look like in practice. Scenario development and system dynamic simulation experiments are used to examine the impact of digitalization on the resilience of the U.S. food sector in the face of catastrophic electricity loss. The study stimulates a critical discourse on the interplay between digital transformation and social resilience and provides insights for the responsible embedding of digital technologies.

Taken together, these studies form the basis for a responsible DSR ecosystem. However, this dissertation recognizes the participatory and emergent nature of such development efforts and can therefore only serve as a starting point. Continued discourse and deliberate action are needed to advance responsible DSR and contribute to the flourishing of our societies in the digital age.

Keywords

Design Science Research, Responsible Innovation, Critical Pragmatism, Grand Challenges, Information Systems, Responsible DSR Ecosystem

Acknowledgement

First, I would like to thank the members of the examination committee, Prof. Dr. Christoph Rosenkranz, Prof. Dr. Detlef Schoder, and Prof. Dr. Stefan Seidel, for their insightful review of this dissertation and their probing questions during the oral defense. I am deeply grateful for their support of my doctoral dissertation.

Then, I would like to express my deepest gratitude to Johanna, whose unwavering support and patience played an indispensable role in my academic journey. Her willingness to engage with my questions and provide insightful perspectives enriched my understanding and sustained me through the arduous process of completing this dissertation. I know it has not always been easy for either of us, but your support has been a constant source of encouragement. So, thank you!

I am deeply grateful to my family, especially Evelyn, Leo, Anna, and Martin, for their constant support and encouragement. Without them, I would not have been able to embark on such an audacious project. Their belief in my abilities and their understanding during the various challenges of this scholarly pursuit have been invaluable.

I am indebted to the colleagues and friends who are also engaged in scholarly pursuits and who have taken the time to engage with my work. In particular, I would like to thank Christoph Rosenkranz for being an open-minded supervisor and mentor who was always happy to provide constructive feedback and encouragement when needed, and all the members of his team, especially Phil, Mario, Niko, Robert and Tim for the inspiring time we spent together discussing research ideas. I would also like to thank Amir Haj-Bolouri, Michael Schlaile, Susan Hanisch, and Dustin Eirdosh for many thoughtful discussions and constant encouragement. Their feedback and constructive criticism were instrumental in refining my ideas and improving the quality of this dissertation and my work in general.

Special thanks are reserved for those individuals who, at various stages of my journey, provided not only emotional support but also encouragement when the outlook seemed daunting. In particular, I would like to mention Sara, Yana, Marie, Max, and Marius. Their belief in my abilities served as a motivating force that fueled my determination to persevere in the face of challenges.

I would like to express my deep appreciation to the scholars in my field who generously participated in the interview study that forms the core of this dissertation. Their insights and willingness to share their expertise contributed significantly to the depth and richness of this research. Without their invaluable contributions, this dissertation would not have taken its current form.

Finally, I would like to acknowledge the collective support of all those who have contributed in various ways to this academic endeavor and whom I have not yet explicitly mentioned. Their help has been indispensable and I am deeply grateful for the role each of them has played in the realization of this dissertation.

. . . .

Alexander Herwix

Contents

Publications

Appendix

List of Tables

List of Figures

Synthesis

Introduction

1.1 Motivation

As humanity has entered the digital age,[1] we are witnessing an accelerating pace of innovation and development in information technology (IT), such as increasingly powerful deep learning-based large language models (e.g., OpenAI, 2023) or other artificial intelligence (AI) technologies and systems that are transforming the world (Brynjolfsson & McAfee, 2012). On the one hand, these advances enable powerful new capabilities and user experiences that can be channeled into unprecedented value propositions and business models, transforming the economic makeup of our societies (e.g., Baskerville et al., 2020; Hanelt et al., 2021; Osterwalder & Pigneur, 2010; Wessel, Baiyere, et al., 2021; Yoo et al., 2010). On the other hand, as these technologies grow in power, their continued development, deployment, and use can also have serious (unintended) consequences and pose severe threats to the well-being of our societies, changing the risk landscape we face (e.g., Center for AI Safety, 2023; Consilience Papers, 2022; Ord, 2020). Thus, the digital age is a double-edged sword, ushering in an era of not only increasing power and capability, but also unprecedented risk and uncertainty. If we are to thrive in the future, it must be our imperative to learn to wield this sword *responsibly* (Jonas, 1984). As the saying goes: "With Great Power Comes Great Responsibility" (2023).

In particular, there is no denying that we need to continue to innovate, and ideally even accelerate innovation, in order to address the grand challenges facing our societies, such as those expressed in the Sustainable Development Goals (Griggs et al., 2013). However, the role of innovation, and especially IT-based innovation, in this context is ambivalent. While it is certainly necessary to help advance humanity's transition away from unsustainable ways of life, it can also stand in the way of such a transformation and even pose significant risks of harm (Vinuesa et al., 2020). Moreover, as the IT we use to innovate becomes more powerful, the potential consequences of the risks of innovation also increase.

For example, as Russell (2019), an eminent scholar in the field of AI research, has warned, AI systems that cross the threshold of becoming as generally intelligent as humans (i.e., reaching artificial general intelligence; AGI) pose an existential threat to humanity if their interests are not aligned with ours, but we currently have no proven plans for safely aligning such AI technologies with our interests and values, nor any way of knowing how close we are to crossing the threshold to AGI.[2] Yet the promise of more powerful and more general AI systems is so great that more and more resources are being invested in their development. This situation has prompted hundreds of prominent AI researchers and other notable figures to sign a public statement that "mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war" (Center for AI Safety, 2023). Moreover, as Harari et al. (2023) cautioned,

[1] The term "digital age" is used in this dissertation to refer to the historical period of increasingly pervasive use of information technology that began in the mid-20th century (Castells, 2011).

[2] This is only an outline of an argument that is developed in much more detail and from a variety of perspectives, for example by Bengio (2023), Bostrom (2014), Christian (2020), Cotra (2022), or Yudkowsky (2008, 2022).

even if one believes that AGI is still a long way off and does not pose an imminent existential threat, even today's AI technologies offer unprecedented generative capabilities that can help to manipulate and control not only individuals, but even the cultural fabric of entire societies, with great potential for abuse and dystopian consequences (see also Consilience Papers, 2022; Seger et al., 2020).

Therefore, responsible stewardship of innovation is of paramount importance in the digital age. A key concept in this regard is *responsible innovation* (RI), which means "taking care of the future through responsive stewardship of science and innovation in the present" (Stilgoe et al., 2013, p. 1570) and involves finding "ways to analyze, describe and change how systems of innovation engage, not just with their intended or envisaged futures, but with a full range of implications" (Owen, Bessant, et al., 2013, p. xii).[3] However, our current understanding of how to implement this societal imperative is still in its early stages (e.g., Blok, 2023; Novitzky et al., 2020; Tabarés et al., 2022). Developing and implementing approaches to RI that leverage the benefits and mitigate the risks of emerging IT is a complex task, particularly given the rapid advances in digital and AI technologies and their far reaching implications (e.g., Buhmann & Fieseler, 2021; Consilience Papers, 2022; Jirotka et al., 2017; Steen et al., 2021).

Since Collingridge's (1980) seminal work on the *dilemma of control*, we already know that "attempting to control a technology is difficult, and not rarely impossible, because during its early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow" (p. 19). Furthermore, even if this dilemma of control can be appropriately managed, questions about the right purposes and motivations of innovations still need to be answered. RI "compels us to reflect on what sort of future(s) we want science and technology to bring into the world, what futures we care about, what challenges we want these to meet, what values these are anchored in, and whether the negotiations of such technologically enabled futures are democratic. It asks how the targets for innovation can be identified in an ethical, inclusive, and equitable manner" (Owen, Stilgoe, et al., 2013, pp. 34–35).

Thus, implementing RI in the digital age presents the information systems (IS) research community with formidable challenges, but also with significant opportunities to conduct relevant and rigorous work that can support the flourishing of our societies (e.g., Davison et al., 2023; Majchrzak et al., 2016; Stahl, 2012; Walsham, 2012). A key development in this direction is the emergence of *design science research* (DSR) as a research paradigm that explicitly considers normativity, transformation, and real-world impact as core dimensions of IS research (e.g., Herwix et al., 2022; Hevner et al., 2004; Myers & Venable, 2014; Vaishnavi & Kuechler, 2015). It does so by going beyond the traditional purview of behavioral science research (BSR), which is to develop and justify theories that explain and predict phenomena, by constructing and evaluating innovative artifacts[4] (and prescriptions

[3] As such, responsible innovation implies a forward-looking and collective understanding of responsibility that recognizes the increasingly global and potentially compounding and long-lasting effects of our local actions (e.g., Stilgoe et al., 2013). As citizens of industrialized societies, we share some responsibility for a global economic system that causes climate change and violates planetary boundaries. Forward-looking responsibility in this context asks us to "take" responsibility and do what reasonably can be done to rein in this out-of-control system. This understanding contrasts with more traditional, backward-looking views of responsibility, which are primarily concerned with assigning responsibility (and blame) for the local and immediate effects of individual actions (Jonas, 1984; Ulrich, 1994).

[4] The term artifact is used in this dissertation in the spirit of Simon (1996, pp. 6–7), who viewed artifacts as those systems or things that are specifically adapted to serve human purposes. For example, a piece of wood becomes an artifact when it is turned into a spear. In the context of IS research, several types of artifacts have been distinguished such as IT artifacts, IS artifacts, information artifacts, technology artifacts, and social artifacts (e.g., Lee et al., 2015; Orlikowski & Iacono, 2001).

for their design) that directly address specific real-world problems or opportunities (e.g., Baskerville et al., 2018; Gregor & Hevner, 2013; Hevner et al., 2004). As such, DSR promises to be an excellent research paradigm not only for studying RI, but also for implementing it.

However, it is clear that still more needs to be done to fully realize the opportunities that exist in this regard (e.g., Davison et al., 2023; Gable, 2020; Galletta et al., 2019; King & Kraemer, 2019; Malhotra et al., 2013; Moeini et al., 2019; Nunamaker et al., 2015; Nunamaker et al., 2017; Ram & Goes, 2021; Walsham, 2012). Specifically, a frequently raised point is that we should become better at channeling the diverse perspectives, methods, and insights from the IS community and related fields into strategic (Gable, 2020), inter-, multi-, or transdisciplinary[5] research efforts (Galliers, 2003; Kroeze et al., 2019; Nunamaker et al., 2013; Nunamaker et al., 2017) that help us responsibly address the grand challenges of the digital age (Jirotka et al., 2017; Ram & Goes, 2021; Winter & Butler, 2011). For example, Winter and Butler (2011) argued:

> The limiting factor on the IS community's legitimacy and impact may be its collective ability (or inability) to form substantial communities of inquiry in a timely fashion. Articulating big, meaningful problems in ways that attract sustained attention from diverse groups of scholars is just as important for advancing a field as conducting studies that resolve and eliminate questions. Without grand challenges and the associated communities of inquiry, the knowledge and collective of ability of a discipline is doomed to remain latent, and largely invisible to broader intellectual and social community around it. (Winter & Butler, 2011, p. 107)

A relevant development in this regard has been the recognition that DSR can serve as a "super-methodology" (Nunamaker et al., 1990) that can help to coordinate even very large-scale strategic, inter-, multi-, and transdisciplinary research efforts that integrate all forms of research into systematic programs of research (Nunamaker et al., 2013; Nunamaker et al., 2017; Ram & Goes, 2021). In particular, as already suggested by Nunamaker et al. (1990) and later demonstrated more rigorously by W. Kuechler et al. (2005), the general understanding of how DSR efforts unfold applies not only to individual research projects, but also to larger streams of research on a common theme, "even when [the work is] conducted by nominally distinct groups (i.e. computer science cf. information systems; academics cf. practitioners)" (W. Kuechler et al., 2005, p. 123).

In practice, this means that DSR need not be limited to organizing individual research projects, but can also help to better coordinate larger, distributed research efforts. For example, Nunamaker et al. (2013) developed and illustrated a DSR framework aimed at facilitating research programs that integrate all forms of IS research in pursuit of real-world impact. Ginige et al. (2018) described how they used a DSR framework to coordinate a complex, multidisciplinary research program to develop a solution to help farmers in Sri Lanka make better-informed decisions about which crops to plant. Paradice et al. (2019) illustrated how a grand challenge framework can be used to stimulate and coordinate a stream of research from nominally distinct groups, even over time frames of several decades. Despite these promising starting points, however, there has been little

[5]In the context of this dissertation, the term interdisciplinary research refers to research in which a topic at the intersection of two or more disciplines is worked on collectively; the term multidisciplinary research refers to research in which a specific topic (possibly outside the general interest of each individual discipline) is viewed and worked on from different disciplinary perspectives; and the term transdisciplinary research refers to research in which disciplinary boundaries are transcended and topics are worked on based on a collectively defined agenda (Kroeze et al., 2019).

progress in IS research as a whole in taking these nascent developments to their conclusion. The mainstream in IS research has yet to recognize the potential of DSR to serve as a "super-methodology". Calls for more strategic, inter-, multi-, and transdisciplinary programs of research highlight the opportunities that await but remain largely unmet (e.g., Desouza & Dawson, 2023; Gable, 2020; Nunamaker et al., 2017; Ram & Goes, 2021). More work is needed to realize this latent potential and to make progress toward developing a more responsible and relevant, as well as more consilient and cumulative tradition.

Against this background, the motivation for this dissertation is to unleash the latent potential of the IS community by initiating and supporting the development of a *responsible DSR ecosystem*[6] that can help implement and advance RI by addressing the grand challenges of the digital age in a responsible and productive manner. As a first step, this means developing a framework for responsible DSR as a conceptual foundation for an open research ecosystem that can bridge the current division in the IS research community between the dominant BSR and the emerging DSR paradigms (e.g., Drechsler & Hevner, 2022; Hevner et al., 2004; Seidel & Watson, 2020; Vaishnavi & Kuechler, 2015) by facilitating the integration of different paradigmatic perspectives. In particular, responsible DSR envisions an open-minded and open-ended pragmatist approach to research in which inquiry is not limited by preconceived notions of appropriate research methods and outcomes (e.g., Dewey, 1938; Ulrich, 2003; Ulrich, 2006), but is guided by the contextual demands of RI in particular situations (e.g., Jirotka et al., 2017; Owen, Bessant, et al., 2013; Stirling, 2006; Ulrich, 1994). Thus, while this dissertation is fortunate to be able to build on the IS community's rich literature and personal experiences with DSR, it nevertheless argues that more work was (and is) needed to break DSR out of the narrow box in which it has traditionally been confined (e.g., W. Kuechler & Vaishnavi, 2008; Levy & Hirschheim, 2012; Nunamaker et al., 2013) and to develop a responsible DSR ecosystem that can responsibly and productively address the grand challenges of the digital age.

1.2 Research Objectives, Questions, and Logic

As motivated in the previous subsection, the overall purpose of this dissertation is to *initiate and support the development of a responsible DSR ecosystem that can help implement and advance RI by addressing the grand challenges of the digital age in a responsible and productive manner*. This purpose requires us to expand our understanding of DSR, broadening its scope from a focus on individual research projects to a "super-methodology" for coordinating an entire research ecosystem. It also calls on us to integrate considerations of RI more deeply into our work. Because this is a complex goal that cannot be achieved in a straightforward manner, it is broken down into three distinct but interrelated inquiries, each of which can move us a bit closer to the realization of the overarching purpose:

- *Inquiry 1* begins with a review of ethics as a key field of reference and motivation for this work. Ethics, understood as the question of what is right and wrong, is a core consideration at the heart of all of science (Churchman, 1995; Ulrich, 2006), but is even more relevant when the goal is RI (Stilgoe et al., 2013; Voegtlin & Scherer,

[6] I speak of a "responsible DSR ecosystem" because the work presented in this dissertation builds on and extends work on DSR. In particular, it aims to adapt DSR to become an appropriate framework for implementing RI by broadening its scope from DSR as an approach for individual research projects to DSR as a "super-methodology" for coordinating an entire research ecosystem. In this context, the term ecosystem is used to emphasize that responsible DSR is a collective effort that depends on the coordination of a variety of parties, such as funders, research participants, editors, and reviewers, so that they can work together in evolving configurations (Stahl, 2022).

2017; von Schomberg, 2013). Therefore, the goal of this review is to provide an ethical foundation for this dissertation by seeking an answer to the research question: *What are key ethical considerations for IS research and how can we deal with them?* Specifically, it integrates reflections from a panel discussion with findings from a broad, narrative literature review (Boell & Cecez-Kecmanovic, 2014a, 2014b) to synthesize different perspectives on ethics in IS research and DSR in particular, and to provide recommendations on how to promote ethical awareness and responsible behavior. In doing so, it advances the ongoing discourse on ethics in IS research with an integrative account of the real-world significance of ethics, and sets the stage for a more systematic and deliberate engagement with RI in IS research and DSR in particular.

- *Inquiry 2* is the main part of this dissertation, which uses a multi-grounded theory development approach (Goldkuhl & Cronholm, 2010, 2018) based on 41 interviews with diverse members of the IS community and scholarly literature from several academic fields to develop a conceptual framework for responsible DSR and make recommendations for its implementation. It addresses the research question: *What are key challenges for responsible DSR in practice, and what recommendations can we make to improve the status quo?* The result is a holistic framework that is centered around four fundamental paradigmatic challenges: ontological, epistemological, axiological, and methodological. By delineating paradigmatic challenges rather than prescribing a specific metaphysical position, the framework is particularly adept at functioning as a boundary object (Star & Griesemer, 1989), thereby providing an open and integrative conceptual foundation that can support the development of a responsible DSR ecosystem.

- *Inquiry 3* concludes this dissertation by initiating an applied responsible DSR project to kickstart the development of a responsible DSR ecosystem. Drawing on scenario development (Goodwin & Wright, 2010; Wright & Goodwin, 2009) and system dynamics based simulation experiments (Dong, 2022; Fang et al., 2018), this inquiry investigates the research question: *How do different degrees of digitalization in society affect the resilience of the U.S. food system in catastrophic electricity loss scenarios?* Based on the simulation results, the concept of digital fragility is developed to explain how the digital transformation of our societies may have the unintended consequence of increasing the systemic risks of catastrophic outcomes. In doing so, it provides a starting point for a critical discourse on the interplay between digital transformation and societal resilience, and possible actions we can take to ensure a responsible embedding of digital technologies in our societies.

Taken together, the three inquiries establish a foundation and vision for a responsible DSR ecosystem. I acknowledge from the outset, however, that this dissertation can only sketch out the first steps in the realization of such a vision. Responsible DSR, as envisioned, is by its very nature a social, participatory, and emergent endeavor that cannot and should not be fully predetermined. As von Schomberg (2013, p. 63) pointed out, RI requires "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to 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)." It is therefore important to emphasize the dialogical orientation and interactive nature of the work ahead (e.g., Shilton et al., 2021). This work can only be a starting point for further discourse and deliberate action.

1.3 Publications of this Dissertation

This cumulative dissertation consists of three publications that reflect the logic described above. Table 1 summarizes the publications with their research question, research approach, contribution, and publication status. As some of the publications were written in collaboration with other authors, Table 1 also details the contributor roles of the author of this dissertation with CRediT (Allen et al., 2019) statements.

Table 1: Overview of the Publications of this Cumulative Dissertation.

Number	**1**	**2**	**3**
Title	*Ethics in Information Systems and Design Science Research: Five Perspectives*	*Threading the Needle in the Digital Age: Four Paradigmatic Challenges for Responsible Design Science Research*	*What Happens When the Machines Stop? Uncovering the Risk of Digital Fragility as the Achilles' Heel of the Digital Transformation of Societies*
Authors	Herwix, A., Rossi, M., Purao, S., Haj-Bolouri, A., Tremblay, M. C., Gregor, S.	Herwix, A.	Herwix, A., Tieman, R., Rivers, M., Rosenkranz, C., Denkenberger, D.
Research Question	What are key ethical considerations for IS research and how can we deal with them?	What are key challenges for responsible design science research in practice, and what recommendations can we make to improve the status quo?	How do different degrees of digitalization in society affect the resilience of the U.S. food system in catastrophic electricity loss scenarios?
Research Approach	Panel discussion with experienced DSR scholars and a broad, narrative literature review (Boell & Cecez-Kecmanovic, 2014a, 2014b)	Multi-grounded theory development (Goldkuhl & Cronholm, 2010, 2018) based on 41 interviews with stakeholders in the DSR community and literature from a variety of academic fields	Scenario development exercise (Goodwin & Wright, 2010; Wright & Goodwin, 2009) and system dynamics based simulation experiments (Dong, 2022; Fang et al., 2018)

Contribution	Summarizes different perspectives on ethics in IS research and DSR and makes recommendations for how to support ethical awareness and responsible behavior.	Develops a holistic paradigmatic framework for responsible DSR and makes recommendations for its implementation.	Provides evidence for the negative impact of higher degrees of digitalization on the resilience of the U.S. food system. Develops the concept of digital fragility to explain how the digital transformation of our societies can increase systemic risks of catastrophic outcomes.
Pub. Status	Published in 2022 at the *Communications of the Association for Information Systems.*	Published as a working paper in 2023. Earlier version presented in 2019 at the *European Conference on Information Systems.*	Published as a working paper in 2023. Earlier versions submitted to *MIS Quarterly* in 2020 (retracted after invitation to resubmit for 2nd round review) and the *Journal of Information Technology* in 2021/22 (rejected after 2nd round review).
CRediT Statement	Conceptualization (lead), Methodology (lead), Writing – original draft (lead), Writing – review and editing (lead), Visualization (lead), Project admin (lead)	Conceptualization, Methodology, Validation, Investigation, Writing – original draft, Writing – review and editing, Visualization, Project admin	Conceptualization (lead), Methodology (lead), Software (lead), Validation (equal), Formal analysis (lead), Investigation (equal), Data curation (lead), Writing – original draft (lead), Writing – review and editing (lead), Visualization (lead), Project admin (equal)

Theoretical Foundations

2.1 Overview

This dissertation builds on three main areas of research. It seeks to integrate them to initiate and support the development of a responsible DSR ecosystem (see Figure 1). The goal is to help implement and advance RI by addressing the grand challenges of the digital age in a responsible and productive manner.

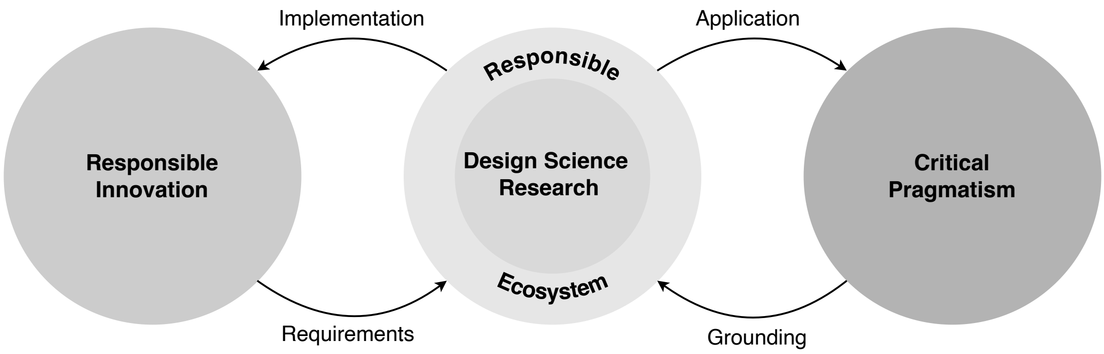

Figure 1: Overview of the Theoretical Foundations of this Dissertation.

First, this dissertation centers on DSR, an emerging approach to IS research known for its emphasis on real-world impact through the development and evaluation of innovative artifacts (e.g., Hevner et al., 2004; Vaishnavi et al., 2019). Specifically, the goal of this dissertation is to establish DSR as an appropriate framework for implementing and advancing RI. This requires expanding its role from an approach for individual projects to a comprehensive "super-methodology" that can help coordinate systematic programs of research (Nunamaker et al., 1990; Nunamaker et al., 2013) and, ultimately, an entire research ecosystem (Morana et al., 2018). This includes all parties relevant to RI, such as funders, research participants, policy makers, editors, and reviewers. I trace the historical development of DSR to place the emergence of this idea in the larger context of IS research.

Second, the goal of a responsible DSR ecosystem is to help implement and advance RI, an ideal and a set of guiding principles for how to manage and govern innovation in a way that is consistent with democratic values (Owen, 2019, p. 27). Thus, I explore the need for RI in the digital age and how it relates to IS research. On the one hand, I identify a great potential of IS research to support RI (e.g., Davison et al., 2023; Majchrzak et al., 2016; Stahl, 2012; Walsham, 2012). On the other hand, I argue that more work is needed to fully realize this potential (e.g., Gable, 2020; Jirotka et al., 2017; Nunamaker et al., 2013; Nunamaker et al., 2017; Ram & Goes, 2021; Winter & Butler, 2011).

Third, the conceptual development in this dissertation is based on *critical pragmatism* as a practical philosophy that provides a sound theoretical foundation for understanding and supporting responsible practice. I introduce the work of John Dewey (1938) and Werner Ulrich (2006) as two key reference points that help to define critical pragmatism for the purposes of this work.

2.2 Historical Development of Design Science Research in Information Systems

The term *Design Science* was coined by Buckminster Fuller (1957), a radical thinker who envisioned *Comprehensive Anticipatory Design Science* as a radical, forward-looking, and holistic approach to problem solving that advocates the application of a scientific mindset to address the greatest challenges facing humanity. In the context of this vision, Fuller emphasized the importance of designing artifacts that meet the needs of individuals and society, considering the broader implications of design decisions and arguing that they should be informed by a deep understanding of the underlying principles and the anticipated consequences of their implementation. Fuller's vision is vividly illustrated in his short book *Operating Manual for Spaceship Earth* (Fuller, 1969), in which he framed our planet as an expertly designed, self-contained, and interconnected system—a "spaceship"—for whose sustainable management we must collectively take responsibility. In particular, he emphasized the need to understand and adapt to the Earth's finite resources and interconnected processes. To do this, Fuller argued that humanity must move from a reactive mindset to a proactive one that anticipates future challenges and develops solutions that are in harmony with the ecological balance of the planet. Thus, in a spirit similar to other visionary contemporaries such as C. West Churchman (1984) or Russel Ackoff (1974), he advocated a holistic approach to systems thinking that considers the global consequences of individual actions.

After Fuller had introduced the term design science, it was quickly picked up by other researchers such as Gregory (1966), who used it to refer to "the study, investigation and accumulation of knowledge about the design process and its constituent operations" (p. 323), or Hubka and Eder (1987), who framed it as the "scientific study of the design activity in its context" (p. 124). In this way, the term became associated with the goal of developing an "explicitly organised, rational and wholly systematic approach to design" (Cross, 1993, p. 66).

In a similar vein, Herbert Simon's (1996) seminal work on *The Sciences of the Artificial* emphasized the pervasiveness of design and its products (i.e., artifacts and artificial systems) in modern life, asserting that "everyone designs who devises courses of action aimed at changing existing situations into preferred ones" (p. 111). Thus, for Simon, all professional activity (e.g., engineering, architecture, business, education, law, medicine, etc.) is a form of design that should also be taught as such. Consequently, he called for the development and teaching of an interdisciplinary "science of design, a body of intellectually tough, analytic, partly formalizable, partly empirical, teachable doctrine about the design process" (p. 113) that would better enable our educational institutions to fulfill their professional responsibilities. For him, science was too concerned with analyzing how things *are* and not enough with how to make things the way they *ought to be* (p. 114). As such, Simon's vision can be viewed as a plea to "improve our understanding of design through 'scientific' (i.e. systematic, reliable) methods of investigation" (Cross, 1993, p. 67). While Simon's work has been widely recognized in many fields, the vision it articulates remains aspirational to this day (You & Hands, 2019).

Amidst this tapestry of visionary ideas, the field of IS research emerged as an arena in which these early works on the relationship between design and science found fertile ground. As summarized by Dickson et al. (1982), the IS field took root at the intersection of computer science and the field of management following the burgeoning use of information technology (IT) for organizational management in the 1960s. The field expanded rapidly in the 1970s, but in the 1980s it began to face challenges stemming from its own growth and ambiguous identity as neither computer science nor management, but something in between. While there was a common sentiment that IS research should be both practically relevant and academically rigorous by leading the charge in intellectual discourse and inquiry regarding the application of IT in business, government, and society (e.g., Keen, 1991), there was much less agreement on how to achieve this (W. Kuechler & Vaishnavi, 2008). While some scholars argued for the adoption of a unifying IS research framework based on common variables and propositions (e.g., Dickson et al., 1982), competing proposals were put forward. For example, Weber (1987) suggested to focus on developing a theory of the behavior of discrete artifacts with longevity as a broader paradigmatic base for IS research. Rather than presuppose a certain set of variables or hypothesis to explain a general phenomenon such as IT adoption, this proposal emphasized the potential benefits of establishing a shared framework for theorizing the contextual behavior of IT artifacts and testing it empirically through design. It is this latter way of thinking that prepared the ground for the emergence of the first steps toward a systematic approach to DSR in IS in the 1990s.

One of the first major steps was taken by Nunamaker et al. (1990), who advocated a more artifact-centered approach to IS research driven by the idea of addressing important real-world problems through multi-methodological research programs that include the development of IT artifacts as an integral part of their methodological portfolio. In particular, they recognized the value of both engineering research,[1] with its focus on developing novel, relevant artifacts, and BSR, with its focus on rigorous description, explanation, and prediction of phenomena, and argued for their integration. To this end, they proposed a general, integrative, iterative, and multi-methodological *Process for Systems Development Research* that integrates *systems development* along with *theory building*, *observation*, *experimentation* into an overarching organizing framework (or "super-methodology," p. 98) for artifact-centered research programs. This process consisted of five stages:

1. *Construct a conceptual framework* (to justify the relevance of the research problem),
2. *Develop a system architecture* (to map out the systems building process),
3. *Analyze and design the system* (to define the design specifications for the implementation of the system),
4. *Build the system* (to demonstrate viability and enable empirical testing),
5. *Experiment, observe, and evaluate the system* (to test its performance and usability as well as observe its impacts on individuals, groups, or organizations).

They supported their argument for the appropriateness of such an approach by describing how several prominent research streams and programs resonated with the methodological structure they outlined.

This was followed by Walls et al.'s (1992) pioneering work on *IS design theories* as vehicles for organizing and sharing prescriptive knowledge about how to achieve particular goals through the design of IT artifacts. As they argued, IS design theories should integrate

[1] Engineering in general is defined as "the application of science to realistic systems which benefit humankind" (Thiel, 2014, p. 2). Engineering research is then used as an umbrella term to refer to all academic research that focuses on engineering. As such, engineering research is part of "the sciences of the artificial," an even broader umbrella term coined by Simon (1996) to refer to all academic fields concerned with design—the development of courses of action aimed at transforming existing situations into preferred ones (p. 111). Given this understanding, DSR in IS can reasonably be viewed as an idiosyncratic subfield within the sciences of the artificial, closely related to and arguably also a form of engineering research.

insights from explanatory (i.e., what is), predictive (i.e., what will be), and normative (i.e., what should be) theories into a structured format that facilitates their practical application in the design of theoretically grounded IT artifacts. In addition, they maintained that IS design theories, like other theories, need to be subjected to empirical refutation to gain credibility, thus supporting the addition of IT artifact development as an important methodological component to the IS research portfolio. Walls et al. (1992) illustrated their conceptual ideas by integrating previous research on executive IS to articulate an IS design theory for what they called "vigilant IS." However, they did not empirically test their IS design theory through artifact development and evaluation, leaving this as a task for future research.

A third contribution was made by March and Smith (1995), who proposed that IS research should be viewed as consisting of two distinct but valid and complementary approaches to science: *design science* (i.e., DSR) and *natural science* (i.e., BSR).[2] For them, the main distinction between these two approaches was *intent*. The design science approach was characterized by the desire to achieve human goals through the construction and evaluation of practical and effective artifacts, while the natural science approach focused on theorizing and justifying theories to explain and predict phenomena. They suggested that these approaches are complementary and could help to realize an artifact-centered approach to IS research in the spirit of Weber's (1987) proposal to develop a theory of artifacts mentioned above. In particular, they asserted that the artifacts built and evaluated with the design science approach could be productively studied with the natural science approach, whose theories could in turn be put to the test in the development of better artifacts. To clarify the nature of such artifacts, they also introduced a taxonomy of research outputs (p. 256-258), which distinguished *constructs* (i.e., the vocabulary of a domain), *models* (i.e., a set of propositions or statements expressing relationships among constructs), *methods* (i.e., a set of steps used to perform a task), and *instantiations* (i.e., the realization of an artifact in its environment). Interestingly, they explicitly rejected the idea that the design science approach should develop and test theories, which was considered to be of interest only to the natural science approach (p. 253-254).

It should be noted that these early developments likely developed independently of each other as none of the works substantially built on or even referred to another. For example, Nunamaker et al. (1990) aimed for a methodological contribution, Walls et al. (1992) focused on epistemological concerns, and March and Smith (1995) attempted to clarify the relationship between the emerging DSR and the more established BSR approach. Thus, while interest in DSR was clearly growing, there was still a lack of an integrated discussion and understanding of what DSR entailed and how it should be done. In particular, DSR was still a niche phenomenon in the most prestigious IS journals, which were dominated by various forms of BSR. This situation set the stage for a phase of consolidation and subsequent rapid development of DSR in IS in the early 21st century.

Arguably the most important step at that time was taken by Hevner et al. (2004), who integrated previous work on DSR into a seminal paper published in MIS Quarterly, one of the most prestigious journals in the IS field. Following March and Smith (1995), they juxtaposed DSR with BSR in an attempt to gain recognition and acceptance for what they called the *design science paradigm*. Central to their work was an IS research framework that conceptualized IS research as being the nexus between the academic knowledge base on the one hand and the practice environment on the other. Only by bringing together

[2]Note that I am using here the original nomenclature of March and Smith (1995), who, following Simon (1996), distinguished between design sciences (i.e., the sciences of the artificial) and the natural sciences. While it may be debatable whether their terminology was appropriate, their distinction effectively prefigured the DSR vs. BSR divide that we still observe today.

both of these sides can IS research achieve the dual demands of relevance and rigor that have motivated it since its emergence (e.g., Keen, 1991). While DSR aims to achieve this "through the building and evaluation of artifacts to meet the identified business need" (p. 79-80), BSR aims to do this "through the development and justification of theories that explain or predict phenomena related to the identified business need" (p. 79). In their view, DSR and BSR are distinct in intent but intertwined in application: "The goal of behavioral science research is truth. The goal of design science research is utility. [... Yet,] truth and utility are inseparable. Truth informs design and utility informs theory" (p. 80). Thus, Hevner et al. (2004) conceptualized a relationship between DSR and BSR that was very similar to that of March and Smith (1995). However, unlike March and Smith (1995), they implicitly acknowledged that design theory in the sense of Walls et al. (1992) could be a possible DSR contribution (p. 87), thereby arguably introducing some incoherence into their own account of the differences between DSR and BSR.[3]

Going beyond their conceptual work on the relationship between DSR and BSR, Hevner et al. (2004) also proposed a set of seven guidelines for DSR that more clearly explained what doing DSR in practice should entail. This sparked a rapid growth of interest in DSR, as there were now some authoritative criteria for evaluating and justifying DSR that, if met, promised publication opportunities in the most prestigious IS journals. However, confusion and disagreement about the nature of DSR remained (Deng & Ji, 2018; Purao et al., 2008).

Walls et al. (2004), in a reflection paper on the implications of their initial rendition of IS design theory (Walls et al., 1992), highlighted how their vision of IS design theory was complementary to, but somewhat neglected by, the IS research framework proposed by Hevner et al. (2004). For them, the role of IS design theory remained underdeveloped in Hevner et al.'s (2004) IS research framework because it did not clarify whether, and if so, how DSR could engage in theorizing—an activity explicitly and exclusively associated with BSR in their framework. The relevance of this tension was only reinforced when Gregor (2006) and Gregor and Jones (2007) further solidified the notion of IS design theory in the IS research discourse.

Over time, this tension has been somewhat alleviated, as scholars have overwhelmingly accepted the idea that theorizing can indeed be part of DSR, but still diverge in their interpretation of what the outcome should be (e.g., Baskerville & Pries-Heje, 2010; B. Kuechler & Vaishnavi, 2008; W. Kuechler & Vaishnavi, 2012; Niehaves & Ortbach, 2016; Venable, 2006). For example, Baskerville and Pries-Heje (2010) argued that the initial proposals for IS design theory by Walls et al. (1992) and Gregor and Jones (2007) were too complex because they unnecessarily combined explanatory and prescriptive components. In their view, IS researchers would be better served by a clear distinction between *explanatory design theory*, which relates general requirements to general components to explain *why* a component is part of an artifact (thus reflecting more traditional BSR theories), and *design practice theory*, which prescribes *how to* construct an artifact.

As of today, while there is still no clear consensus on how best to capture and accumulate design-related knowledge (e.g., Goldkuhl & Sjöström, 2021; Gregor et al., 2020; Lukyanenko & Parsons, 2020; Rohde et al., 2017; vom Brocke et al., 2020), there is a general agreement that theorizing can (and, for some, should; see Baskerville et al., 2011) indeed be a central aspect of DSR. This has led to the understanding that DSR contributions can be made at different levels of abstraction, for example, ranging from practical contributions related to instantiations (e.g., existence proofs), to nascent design

[3]If DSR is defined by building and evaluating useful artifacts, and BSR is defined by developing and justifying theories, the recognition that DSR also develops and justifies theories seems to call into question theory as a distinguishing feature.

knowledge (e.g., constructs, models, frameworks, architectures, and design principles), to well-developed design knowledge in the form of comprehensive design theories (see, e.g., Baskerville et al., 2018; Gregor & Hevner, 2013; vom Brocke et al., 2020). At all of these levels, DSR contributions must demonstrate an advance over the current state of the art in routine design, in the form of insight into a new solution to a known problem (i.e., improvement), an extension of a known solution to a new problem (i.e., exaptation), or a new solution to a new problem (i.e., innovation) (Gregor & Hevner, 2013). In addition, some scholars have even recognized the possibility that DSR may directly contribute to the development of BSR theories and knowledge (e.g., Baskerville et al., 2015; Briggs & Schwabe, 2011), leading to calls for broadening the scope of DSR to become an overarching and integrative approach to IS research (e.g., Briggs & Schwabe, 2011; Nunamaker & Briggs, 2011; Nunamaker et al., 2013; Nunamaker et al., 2017).

Another source of confusion that remained after Hevner et al.'s (2004) publication was the methodological structure of DSR. While the methodological contribution by Nunamaker et al. (1990) still provided a useful template for structuring multi-methodological research programs involving the development of IT artifacts, the seminal work by Hevner et al. (2004) prompted a number of scholars to revisit the issue of DSR's methodological structure and propose updates to existing frameworks, prominent examples being Vaishnavi and Kuechler (2005), Peffers et al. (2007), and later Sein et al. (2011).

Vaishnavi and Kuechler (2005) and Peffers et al. (2007), in the spirit of Nunamaker et al. (1990), promoted methodological frameworks centered on artifact development processes inspired by already more established areas of design and engineering research such as computer-aided design and software engineering. Vaishnavi and Kuechler (2005) distinguished five iterative but linear phases:

1. *Awareness of Problem* (produces a research proposal),
2. *Suggestion* (produces a tentative artifact design),
3. *Development* (produces an instantiated artifact),
4. *Evaluation* (produces performance measures), and
5. *Conclusion* (is reached when "good enough" results are achieved).

In a similar vein, Peffers et al. (2007) differentiated six iterative but linear phases:

1. *Problem Identification and Motivation* (identifies a relevant and significant problem),
2. *Define Objectives for a Solution* (establishes clear objectives for designing a solution to the identified problem),
3. *Design and Development* (creates a novel artifact),
4. *Demonstration* (shows the functionality and effectiveness of the artifact),
5. *Evaluation* (assesses the artifact's utility, quality, and impact), and
6. *Communication* (documents and shares the findings and outcomes of the DSR process to relevant audiences).

As can be seen by comparing the outlines of these methodological frameworks, they advocated structurally very similar approaches to DSR, differing more in emphasis and presentation than in substance. They also bear a strong resemblance to the Process for Systems Development Research proposed by Nunamaker et al. (1990). However, in contrast to Nunamaker et al. (1990), they were much less explicit about the possibility of multi-methodological research programs using DSR as a "super-methodology". Following the sharp distinction between DSR and BSR made by Hevner et al. (2004), it seems that their primary interest was in strengthening the position of DSR as a viable research approach for individual IS research projects, rather than as an overarching and integrative research approach advocating an organizing logic for multi-methodological IS research programs in general (cf. Briggs & Schwabe, 2011; Nunamaker & Briggs, 2011; Nunamaker et al., 2013).

Similar in their focus on strengthening the position of DSR as a viable research approach for individual IS research projects, but different in their methodological genealogy, Sein et al. (2011) proposed *Action Design Research* (ADR) as a softer, more interactive, intervention-oriented approach to DSR influenced by *action research* (Baskerville & Wood-Harper, 1998). While engineering-inspired DSR approaches such as those of Vaishnavi and Kuechler (2005) and Peffers et al. (2007) can be excellent for addressing relatively general problems with stable contexts, Sein et al. (2011) argue that such approaches can lead researchers interacting with practice to neglect "the role of organizational context in shaping the design as well as [...] the deployed artifact" (p. 38). In their view, artifacts should be viewed as ensembles of material and organizational features that become entangled during development and use (Orlikowski & Iacono, 2001), blurring the sharp distinction between artifact development and use generally assumed in more engineering-inspired DSR. Given this understanding, the separation of the *building* of an artifact from the *intervention* that puts it into practice and the *evaluations* that test its impact, as advocated in engineering-inspired DSR, can become problematic because the emergent nature of the ensemble artifact may not be adequately recognized. To address this problem, Sein et al. (2011) proposed ADR as a DSR approach in which *Building, Intervention, Evaluation* (BIE) is recognized as a single, inseparable stage among three other interleaving stages, and for which they articulated a total of seven principles to guide and support its implementation:

1. *Problem Formulation* (should aim for *Practice-inspired Research* and a *Theory-ingrained Artifact*),
2. *Building, Intervention, Evaluation* (should recognize *Reciprocal Shaping*, and aim for *Mutually Influential Roles* as well as *Authentic and Concurrent Evaluation*),
3. *Reflection and Learning* (should support *Guided Emergence*),
4. *Formalization of Learning* (should produce *Generalized Outcomes*).

Proposing an integrative BIE stage reflected a broader shift in practice toward more *agile* forms of IS development that advocate incremental design, and a highly iterative, as well as collaborative, approach to IS development aimed at continuous customer involvement to enable more adaptive and innovative organizations (Baham & Hirschheim, 2022). In particular, agile IS development approaches tend to emphasize deliberately short and integrated design, development, and evaluation cycles, so called *sprints*, that orchestrate all the activities required for new product releases (Meyer, 2018). If such practices should also be applied in DSR projects, the recognition of integrative BIE stages seemed to be a useful methodological advance. However, it has also further blurred the line between DSR and BSR by infusing design-oriented research with elements of action research (a research method historically more closely associated with BSR; Baskerville & Wood-Harper, 1998). Particularly when looking at hybrid methods such as ADR, it becomes clear that DSR can span the entire spectrum of IS research if desired (Baskerville et al., 2015; Briggs & Schwabe, 2011), making it difficult to maintain strict boundaries between DSR and BSR and ultimately raising questions about how to properly frame the relationship between them (Herwix & Rosenkranz, 2018).

Thus, both of the two aforementioned major sources of confusion that have driven research on DSR in recent years (i.e., the role of theorizing in DSR and the methodological structure of DSR) seem to have arisen from an overly narrow perspective on the scope and ideals for DSR as they were articulated by Hevner et al. (2004). Following March and Smith (1995), they argued for a narrow DSR paradigm focused on the engineering-inspired development of IT artifacts that could be clearly distinguished from BSR, but such a view was simply too narrow and limited to accommodate the breadth of contributions (Baskerville et al., 2015; Herwix & Rosenkranz, 2018) and diversity of perspectives (Rai, 2017a), and lift the full potential of DSR (cf. Alter, 2012; Briggs & Schwabe, 2011; W. Kuechler & Vaishnavi, 2008; Levy & Hirschheim, 2012; McKay et al., 2012).

Nevertheless, Hevner et al.'s (2004) framework continues to have a strong influence on the current DSR discourse, so much so that even many of the most recent DSR frameworks still aim to maintain a dualistic relationship between DSR and BSR, where each is complementary to, but distinct from, and ultimately incommensurable with the other (e.g., Drechsler & Hevner, 2022; Seidel & Watson, 2020; Vaishnavi et al., 2019), rather than developing Nunamaker et al.'s (1990) key insight that DSR can serve as a general, overarching "super-methodology" for artifact-centered, multi-methodological IS research programs (see Briggs & Schwabe, 2011; Nunamaker & Briggs, 2011; Nunamaker et al., 2013; Nunamaker et al., 2017). However, it is primarily this latter view that promises to provide a strong foundation for the systematic development of high-impact, strategic, inter-, multi-, or transdisciplinary programs of research (Gable, 2020; Nunamaker et al., 2013; Nunamaker et al., 2017) that can address the grand challenges facing our societies in the digital age in a responsible and productive manner (Jirotka et al., 2017). Thus, work aimed at refocusing the paradigmatic underpinnings of DSR seems particularly promising for advancing IS research by supporting the development of a responsible DSR ecosystem that is better able to unleash the full potential of IS research.

2.3 The Need for Responsible Innovation in the Digital Age

Responsible innovation (RI) is an ideal and a set of guiding principles for how to manage and govern emerging technologies (and also other innovations) in a way that is consistent with democratic values (Owen, 2019, p. 27). Innovation in this context can be broadly understood as the process of introducing useful new artifacts into societies, such as new or recontextualized ideas, practices, technologies, products, or services (Bessant, 2013). RI in particular calls for innovation that is "more anticipatory, more reflexive, more inclusive, deliberative, open and, in total, more responsive" (Owen, 2019, p. 27) and the institutional and systemic transformations needed to achieve this (e.g., Owen et al., 2021; Stahl, 2022). The goal is to bring about "a collective commitment of care for the future through responsive stewardship of science and innovation in the present" (Owen, Stilgoe, et al., 2013, p. 36) and to foster "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)" (von Schomberg, 2013, p. 63). As such, RI should not be seen as a ready-made solution, but rather as a contextual, forward-looking challenge to come together and take responsibility for creating a better future. It asks us: (1) what future we want to create; and (2) how we can responsibly bring about that desired future, given the inherent uncertainty and unpredictability of the future and our own ignorance and ambivalence about how to shape it (Owen, 2019, p. 29).

This basic understanding of RI builds on decades of academic discourse (e.g., in the fields of science and technology studies, computer ethics, philosophy of technology, and technology assessment), which has extensively discussed vexing questions about how to manage the societal consequences of emerging technological developments. Collingridge (1980), for example, developed the aforementioned *dilemma of control* to capture the difficulty of foreseeing and managing the (unintended and long-term) consequences of innovation (see also Consilience Papers, 2022). Even today, finding ways to address this dilemma remains one of the key challenges for RI, inspiring calls for more open, participatory, diverse, incremental, flexible, and reversible governance processes that acknowledge our fallibility and ignorance in anticipating the societal consequences of innovations (Genus & Stirling, 2018).[4]

With the emergence of potentially disrupting technologies such as nanoscale manufacturing, genome editing, geoengineering, and AI the discourse around RI has grown significantly and also started to involve policymakers directly, who in the early 2010s began to promote the closely related notion of *responsible research and innovation* (RRI) as a guiding framework for the European Commission's Science in Society program (Owen, 2019).[5] It is therefore not surprising that RI itself remains a contested concept, with a diverse and growing community of people seeking to shape its development and implementation (Brundage & Guston, 2019; Timmermans & Blok, 2021). As a result, much of the current RI discourse focuses on the challenges associated with the implementation of RI and calls on us to advance our understanding of how to foster RI in messy, contested, and often constrained real-world practice without compromising the ideals it represents (e.g., Blok, 2023; Owen, 2019; Owen et al., 2021; Shanley et al., 2022; Tabarés et al., 2022). In a sense, RI is like a ship being built at sea that calls for the recursive application of its own ideals and principles to its ongoing development and implementation (cf. Timmermans et al., 2020).

RI is gaining particular importance and recognition in the digital age, where advances in IT, such as those associated with automation, big data, blockchain, and AI promise to impact and transform almost all areas of life (Brynjolfsson & McAfee, 2012; Jirotka et al., 2017). As Jirotka et al. (2017) outline, these developments raise particular challenges for RI, which they discuss as being related to the (1) *processes*, (2) *products*, (3) *purposes*, and (4) *people* associated with RI.

First, the processes of RI are affected by the unprecedented speed of development and deployment of IT. Whereas novel physical technological developments often require sophisticated logistics that can take many months or even years to set up and to be deployed at scale, the Internet enables the rapid development of novel IT artifacts and their deployment to millions or even billions of users within minutes to hours at comparatively

[4]In a similar vein, *Jevons Paradox*—the recognition that efficiency gains in resource use tend to increase rather than decrease resource consumption (Giampietro & Mayumi, 2018)—is another example of how the outcomes of innovations can be counterintuitive and hard to fully anticipate (Blok & Lemmens, 2015).

[5]RI and RRI are discourses so closely related that they are often considered as being significantly intertwined and entangled in a contested discursive web (Randles et al., 2016). Sometimes the hybrid label R(R)I is used to acknowledge this circumstance. For the sake of readability, this paper will use the simpler label RI to refer to the discursive web in aggregation.

little cost. RI processes need to be designed with this step change in mind because loss of control over the proliferation of IT-related artifacts such as software and data is a distinct possibility, which poses new societal risks when considering the impacts associated with, for example, the growth of cybercrime (IC3, 2021) or the spread of misinformation (Seger et al., 2020).

Second, viewed as products of RI, IT-related artifacts also present new challenges because of their inherent logical malleability (Moor, 1985) and interpretive flexibility (Doherty et al., 2006). IT is almost limitless with regard to its application, which means that the outcomes of its development and use are extremely difficult to anticipate. A striking example of this is the field of deep learning-based large language models, where it remains incredibly difficult (if not impossible) to predict what capabilities such AI systems will have. At least for now, we have to estimate their capabilities based on explicit behavioral testing (e.g., OpenAI, 2023), which creates an uncomfortably large space for unintended consequences and unforeseen risks if it turns out that we have missed some crucial details.

Third, in terms of the purpose of RI, the increasingly pervasive embedding and use of IT in all areas of life blurs the boundaries between systems, features, and functionalities, thus, opening spaces for rethinking taken-for-granted assumptions and challenging our sense of orientation (Grunwald, 2007). As the digital age touches, transforms, and disrupts everything, RI is challenged to ensure that the future we seek is desirable not only for those in the driver's seat, but for all passengers on the ride. This challenge is particularly pertinent in light of existing global inequalities that threaten the adequate participation of large segments of the world's population (Walsham, 2017). It has been argued that developing an orientation towards addressing the grand challenges that are demonstrably relevant to all of humanity can be a responsible way forward (Schroeder & Kaplan, 2019).

Fourth, the digital age also brings with it new opportunities and challenges for how people can become implicated in RI. For example, the Internet enables open research and innovation practices that have the potential to productively connect people across the world (von Schomberg, 2019), but it also diffuses responsibility among more actors than ever before, exacerbating the so-called "*problem of many hands*" (Poel et al., 2015), which refers to the difficulty of assigning responsibility for a (bad) outcome if multiple people contributed to it. Preventive measures that foster a culture of forward-looking responsibility, based on a collective commitment to care for the future and a willingness to reflect and deliberate in order to take responsive, anticipatory action in the present, are increasingly needed to help contain the risks posed by this problem (Owen, Stilgoe, et al., 2013; Poel et al., 2015).

Taken together, these developments underscore the growing importance of RI in the digital age. Never before have the stakes for RI been higher and its implementation more urgent (cf. Center for AI Safety, 2023; Consilience Papers, 2022; Ord, 2020). In an increasingly digital, interconnected, and globalized world, implementing RI means addressing the grand challenges of our time (Schroeder & Kaplan, 2019). Doing so requires inclusive, inter-, multi-, and transdisciplinary collaboration and coordination on an unprecedented scale, with an unprecedented level of collective care for a desirable future that works for all (von Schomberg, 2019).

Drawing on Winter and Butler (2011), I note that the ambition of RI itself has all the hallmarks of a (meta) grand challenge. One that aligns perfectly with the interests and expertise of IS research as an interdisciplinary field that has long been concerned with how to successfully design and implement socio-technical IS (Sarker et al., 2019). For example, when RI calls for increased coordination through open research and innovation practices, this requires support from digital platforms and service ecosystems, and IS research has shown a strong interest in understanding how to successfully design, implement, and govern these (e.g., de Reuver et al., 2018; Hein et al., 2020; Sun & Gregor, 2023; Wang, 2021). RI also calls for more ethical methods of technology development, and IS research has a strong history of developing IS development methods (e.g., Mumford, 1995) through method engineering (Goldkuhl & Karlsson, 2020). In addition, IS research has at times demonstrated its ability to focus directly on addressing grand challenges in a responsible manner (e.g., Davison et al., 2023; Majchrzak et al., 2016; Walsham, 2012).

Despite these promising opportunities for IS research to help implement and advance RI, it is clear that more concerted and larger efforts would be desirable to more fully realize the available potential (Stahl, 2012; Winter & Butler, 2011). As already alluded to in the preceding section, much of IS research has been limited by a narrow focus on small-scale research studies in business settings (cf. Nunamaker & Briggs, 2011). Strategic, inter-, multi-, or transdisciplinary programs of high-impact research (Gable, 2020; Nunamaker et al., 2013; Nunamaker et al., 2017) that can systematically address the grand challenges of the digital age are exciting but still too rare exceptions (Desouza & Dawson, 2023; Ram & Goes, 2021). This dissertation aims to address this general situation by initiating and contributing to the development of a responsible DSR ecosystem specifically designed to help implement and advance RI by addressing the grand challenges of the digital age in a responsible and productive manner.

2.4 Critical Pragmatism as a Theoretical Foundation for Responsible Practice

Motivated by the developments outlined in the previous sections, a key aim of this dissertation is to develop a paradigmatic framework for responsible DSR that can underpin and support the emergence and evolution of a responsible DSR ecosystem (cf., Kuhn, 2012). This, however, requires a sound theoretical framing and understanding of what it means to act responsibly. While the RI discourse focuses on parts of the answer by addressing this question from a societal and governance perspective, it can also be approached from a broader philosophical perspective that focuses on responsible practice in general. Such a perspective can further guide, support, and critically challenge the existing RI discourse and ground the development of a theoretical foundation for responsible DSR.

With this goal in mind, this dissertation draws on the work of John Dewey (1938) and Werner Ulrich (2006, 2016) as complementary anchor points for its own development. Both authors contributed to a research program that Ulrich (2006, 2016) called *"Critical Pragmatism"*—the vision of a practical philosophy for responsible practice based on a "critical turn" in pragmatism. While Dewey (1938) presented one of the original, "classical" accounts of the onto-epistemological dimensions of pragmatism that is still regarded as seminal contribution that continues to influence our understanding of scientific practice (Brown, 2012), Ulrich (2006) developed the ethical and methodological dimensions of pragmatism, recognizing more explicitly the "in-built tension between the contextualist and the universalist poles of sound [and ethical] practice" (Ulrich, 2016, p. 11) and offering some practical suggestions for "a methodological renewal and development of pragmatism" (Ulrich, 2016, p. 13) in light of this. Taken together, then, the works of these two authors provide a comprehensive theoretical foundation that offers a sound understanding of responsible practice and fits exceedingly well with the aims of this dissertation.

In what follows, I will first present Dewey's (1938) classic account of pragmatism, then present Ulrich's (2006, 2016) proposal for a critical turn in pragmatism, and finally summarize why critical pragmatism provides a solid foundation for this dissertation.

2.4.1 John Dewey's Classical Account of Pragmatism

This dissertation builds on Dewey's (1938) *Logic: The Theory of Inquiry* because it represents one of the most systematically developed of the classical accounts of pragmatism and remains highly relevant to our current understanding of scientific practice (Brown, 2012). As discussed by Barrett and Aiken (1962), Dewey can be seen as the third member of the pragmatist triumvirate, continuing and refining the work begun by the founders and early proponents of pragmatism, his personal and professional acquaintances, Charles Sanders Peirce and William James:

> It suffices at this point to say that if Peirce may be regarded as the Socrates of pragmatism, and James as its Plato, Dewey is certainly its Aristotle. The philosophical writings of Peirce and James are essentially collections of essays, rich in insight, powerfully written, and immensely suggestive to the sympathetic reader. But they sometimes give the impression of being sketches for a philosophy rather than the finished portrait. It was Dewey's role to fill in the outlines provided by Peirce and James. (Barrett & Aiken, 1962, pp. 49–50)

In his seminal work, Dewey (1938) explores the unfolding of various forms of inquiry. This exploration ultimately leads him to a reconceptualization of logic, wherein logic becomes both a product of and a standard for the process of inquiry (Hildebrand, 2023). As he states, "all logical forms (with their characteristic properties) arise within the operation of inquiry and are concerned with control of inquiry so that it may yield warranted assertions" (Dewey, 1938, pp. 3–4). Thus, Dewey sees logic and inquiry as being in a recursive relationship, with the standards of logic emerging from but also, in turn, constraining the self-correcting process of inquiry (Brown, 2012, pp. 267–268). In this sense, logic is "an attempt to clarify and codify the most general normative constraints that we have discovered as a result of our inquiries, successful and unsuccessful" (Brown, 2012, p. 268). As evident from this brief description, the scope and ambition of Dewey's (1938) work is too vast to be discussed here in detail. Following Brown (2012), we thus limit our discussion to three main concepts that help to summarize the general outline of Dewey's theory of inquiry:

1. *Situations* as both the source and the control of inquiry,
2. *Inquiry* as the process of transforming problematic situations, indeterminate with respect to their issue, into determinate situations where elements form a robustly stable integrated whole, and
3. *Warranted assertibility* of final judgment as the legitimate end and endpoint of inquiry.

First, the starting point for Dewey's theory is that "inquiry always takes place in, is immediately concerned with, and is guided by a situation" (Brown, 2012, p. 268). As such, a situation can be understood to be the stage on which inquiry unfolds, a stage which stretches to cover all aspects of the environment, including the inquiring parties and their perspectives but also all other circumstances and data, that are relevant to an inquiry in question. Taken this way, situations are subjective (i.e., relative to the perspectives and practices of the involved parties) but at the same time objective (i.e., consisting of concrete elements of the natural world), providing the foundation for a sophisticated form of contextualism that advocates a middle path between atomism and holism (Brown, 2012, pp. 268–274).

In particular, while a situation is composed of various identifiable elements, it is also experienced as "a whole in virtue of its immediately pervasive quality" (Dewey, 1938, p. 68). As such, experience is understood to be the core link by which individuals come to engage with and make sense of the various elements within a situation. Experience is not a passive observation, but an active, participatory process that involves continuous interaction and transaction with the environment, resulting in unique, evolving situations. Thus, situations are not static, isolated entities, but dynamic, only vaguely bounded phenomena that shape and are being shaped by experiencing individuals acting in context (cf. Herwix & Zur Heiden, 2022).

For example, consider a scenario where a group of scientists is conducting an IS research study in a large for-profit organization. The situation in this case includes not only the scientists themselves, with their own interests and perspectives, but also the members, socio-technical infrastructure, and supply chain of the organization under study, its (potential) customers, and relevant parts of the politico-economic and the even broader socio-techno-ecological environment. The researchers' perspectives, their research tools, and the data they collect are all intimately connected to and interact with this unique and ever-changing contextual environment. As such, they can adjust their research methods based on new observations, collaborate with local experts to gain a deeper understanding of the organizational ecosystem, and consider the implications of their findings for both the broader environment and the local parts of the organization with which they are directly interacting. In this way, situations are both the source and the control of inquiry, since they call for particular experimental operations and determine their results and effectiveness.

Second, at the heart of Dewey's theory is "the basic conception of inquiry as determination of an indeterminate situation" (Dewey, 1938, p. iii). This conceptualization is based on the understanding that situations are indeterminate (and can become problematic) for the individuals involved insofar as their habitual ways of dealing with them do not reliably produce the intended results, and determinate insofar as inquiry has established their constitutive distinctions and relations in such a way as to bring all of their elements into robustly stable integrated wholes that permit productive action and warrant reaching final judgments (Dewey, 1938, pp. 104–105).

In this sense, inquiry is a *transactional* process because it requires a purposeful interplay of subjective experiences and objective situations, which occurs only when an individual recognizes a situation to be problematic[6] enough to engage in a controlled or directed transformation toward resolution (Brown, 2012, p. 278). This means that the indeterminacy of a situation is only a necessary, not a sufficient, condition for inquiry; that engaging in inquiry involves a component of recognition and choice; and that uncontrolled or undirected action without purpose does not count as inquiry.

In a similar vein, inquiry is also a *transformative* process because it not only changes the inquirer's conceptual understanding of a situation, but also affects the existential, material conditions of a situation through immediate changes in the potentialities of constituent elements and subsequent physical changes that result from putting the newfound understanding into practice (Brown, 2012, p. 279). As such, even inquiry into an abstract mathematical problem transforms the existential, material conditions of a situation by virtue of its consequences.

[6]Note that the term problematic is used here in a functional rather than a normative sense. Situations are recognized to be problematic because habitual ways of acting do not reliably produce the intended results, not for normative reasons. In this sense, recognizing situations as problematic need not be associated with a negative outlook, see, for example, the case of science where the identification of deep and interesting problems is generally understood to be a positive result.

In terms of procedural structure, inquiry is best understood as two-dimensional with (1) a generic, linear sequence of *three temporal stages* (i.e., indeterminate situation, inquiry, and final judgment standing for the settled outcome of inquiry grounded in a resolved situation) that mirrors the aforementioned definition of inquiry, and (2) a generic, nonlinear set of *four functional phases*[7] (i.e., institution of a problem, determination of a problem-solution, reasoning, experimentation) that constitute the pattern of inquiry (see Figure 2; Brown, 2012, p. 280). Whereas the first dimension captures the linear sequence of inquiry as it transforms indeterminate into determinate situations, the second dimension recognizes the complex, nonlinear, and dynamic ways in which inquiry unfolds in any given situation. Taken together, they offer a sophisticated procedural understanding of how progress can be attained.

Central to this view is the recognition that successful inquiry requires moving through four functional phases, or "key steps," whose specific sequence cannot be predetermined, but must be directed with respect to the dynamically evolving situation in focus (Brown, 2012, pp. 280–287; Dewey, 1938, pp. 104–114):

1. The *Institution of a Problem* phase emphasizes that useful and appropriate problem framings do not exist ready-made in nature, but must be constructed as part of the inquiry (see also Andersen et al., 2022; Simon, 1973). The goal of this phase is to consider all relevant details of the indeterminate situation and to relate them in such a way that opportunities for effective resolution become visible.

2. In the *Determination of a Problem-Solution* phase, inquiry focuses on identifying and generating solutions hypotheses that suggest fruitful avenues for moving forward. The goal of this phase is to direct the inquiry toward the most promising directions.

3. The *Reasoning* phase is central to the inquiry and is concerned with the refinement and critical examination of solution hypotheses, which involves relating them to larger theories or frameworks, anticipating the consequences of putting them into practice, and proposing experimental tests based on them. The goal of this phase is to effectively coordinate empirical observations so that solution hypotheses can be tested and refined.

4. In the *Experimentation* phase, solution hypotheses are tested by intervening in the situation and putting them into practice. This can be done in a variety of ways, such as field studies or controlled experiments. The goal of this phase is to produce high quality empirical data that enable a rigorous testing of solution hypotheses and ultimately justify a final judgment as a successful conclusion to the inquiry.

As should be clear from the preceding discussion, feedback loops between the phases are essential to the successful conduct of inquiry. They allow for continuous adjustments and refinements as new data become available and as the understanding of the situation evolves. Inquiry is not a one-way, linear process but a dynamic, iterative cycle that can loop back to any of the previous phases if necessary. For example, problems may need to be reframed and solution hypotheses reconsidered as data from experimentation are

[7] Please note that there is no generally agreed upon set of the functional phases of inquiry as Dewey's own work remains somewhat ambiguous about their demarcation as it evolved over time. As Brown (2012, pp. 280–287) discusses it seems reasonable to distinguish between four to six functional phases. We choose the basic set of four functional phases to stay closely aligned to Dewey's (1938, pp. 104–114) seminal account in his *Logic: The Theory of Inquiry*, where the observation of an indeterminate situation is suggested to be an antecedent and the institution of a problem the starting point of inquiry. Thus, we diverge here from Brown (2012), who decided to include observation as a fifth functional phase of inquiry.

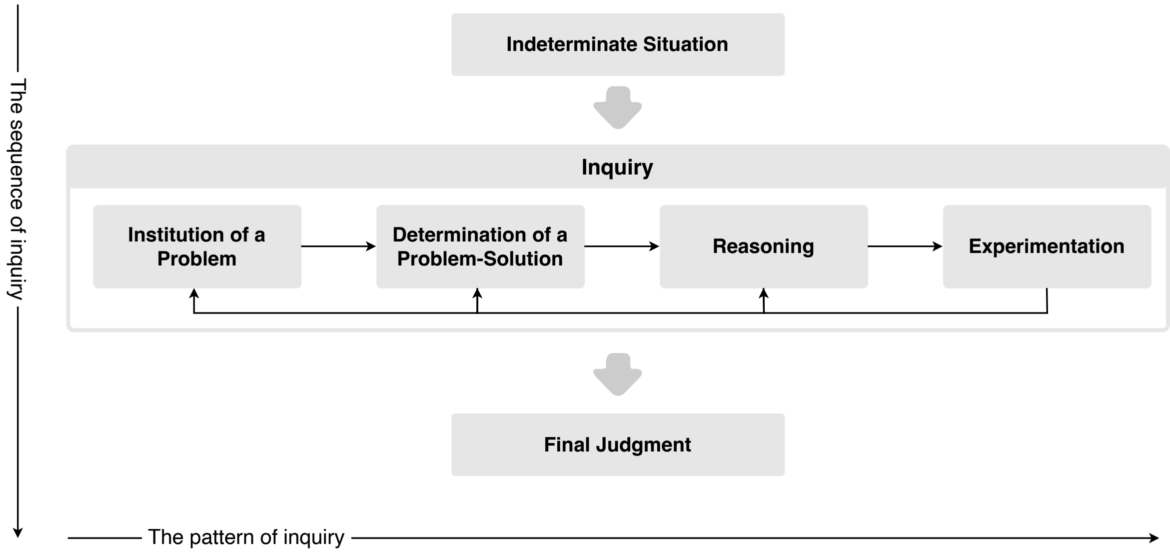

Figure 2: The Two-Dimensional Procedural Structure of Inquiry (based on Figure 1 from Brown, 2012, p. 280).

confronted, leading to an directed and controlled but iterative and nonlinear unfolding of inquiry over time.

Third, the end and legitimate endpoint of inquiry is the warranted assertibility of final judgment. This means that at least one of the solution hypotheses has lent itself to the successful transformation of the indeterminate into a determinate situation, with other solution hypotheses having been more or less ruled out as viable alternatives (Brown, 2012, pp. 280, 297).[8] This establishes a rigorous standard, for it requires that all functional phases of inquiry have been thoroughly carried out with an eye toward achieving "functional fitness" with respect to the situation in focus, and that the final judgment itself "can be used as a settled means to future inquiry" (Brown, 2012, pp. 280, 297).

According to Dewey (1941, p. 169), warranted assertibility thus understood can be seen as "a definition of the nature of knowledge in the honorific sense according to which only true beliefs are knowledge." As such, it defines the conditions under which we can speak of knowledge and truth, namely, only when inquiry has successfully concluded in a settled result in the form of a warranted assertion that can underpin and direct future inquiry, the final judgment. Thus, the logical status of the outcome of successful inquiry (i.e., the final judgment) is distinguished explicitly from the status of the intermediate material (i.e., solution hypotheses) that is inquired into (Dewey, 1938, pp. 118–119). Only the final results produced by means of inquiry should be qualified as knowledge.

However, warranted assertibility is inquiry relative, meaning that "what is warranted on the basis of inquiry in a particular situation may no longer have that status some time down the road, in a very different situation" (Brown, 2012, p. 297). As situations change and understandings evolve, there is simply no guarantee that judgments formed in the past will continue to have force in the present. In response, Dewey's theory proposes a coherent, antiskeptical fallibilism by asserting that while no judgment is perfect or even likely to remain a warranted assertion indefinitely, there must be a positive reason, grounded in a problematic situation, to doubt already settled judgments (Brown, 2012, p. 287).

[8] For this, "it is necessary that data (provided by observation)[, which are used to test solution hypotheses,] be new, or different from those which first suggested the inferential element [. . .]. It is important that they be had under as many different conditions as possible so that data due to differential origins may supplement one another" (Dewey, 1941, p. 173).

Dewey thus rejects any form of dogmatism and recourse to notions of "self-sufficient, self-possessed, and self-evident truth" (Dewey, 1941, p. 172), but rather advocates a melioristic view of the world in which iterative progress and refinement of our conditions is not only possible but is called for (Barrett & Aiken, 1962, pp. 80–81). In this view, while the natural goal of inquiry, and of scientific inquiry in particular, is to develop warranted assertions of a general nature that can be asserted long into the future, the dynamic and evolving nature of the world demands constant and ongoing vigilance in identifying emerging problematic situations and in updating beliefs in the light of new data. As such, inquiries are seen as contributing iterative updates to our understanding that can be leveraged for the ongoing improvement of our collective conditions. In this way, successful inquiry tends to pave the way for more, rather than less, future inquiry (Brown, 2012, p. 279).

In summary, Dewey's theory of inquiry provides a comprehensive framework for understanding the process of inquiry and its role in (scientific) practice. At its core, this theory emphasizes the dynamic interplay between inquiry and the situations in which it unfolds. It stresses that inquiry is deeply embedded in the context, involving both subjective perspectives and objective elements of the natural world. It also highlights the iterative and adaptive nature of inquiry, with its phases of problem institution, problem-solution determination, reasoning, and experimentation subject to dynamic adjustments. Moreover, it sets a rigorous standard for the generation of knowledge through the warranted assertibility of final judgment, while remaining fallibilistic, recognizing that evolving situations may require the revision of previous judgments. In an ever-changing global landscape, it calls for constant vigilance, adaptability, and the pursuit of knowledge as a collective, progressive endeavor.

2.4.2 Werner Ulrich's Call for a Critical Turn of Pragmatism

In recent years, pragmatism has been increasingly recognized as a powerful lens for looking at the world that can help us make sense of how practitioners and academics behave in practice, and thus provides us with an appropriate theoretical foundation to inform our work (e.g., Brown, 2012; Dalsgaard, 2014; Goldkuhl, 2011; Lee & Nickerson, 2010; Ormerod, 2006, 2020). For example, Ormerod (2020), informed by his experience in Operations Research (OR), examines in detail how pragmatism can support professional practice in general, Dalsgaard (2014, p. 148) argued that "pragmatism can serve as conceptual scaffolding for developing the discourse of design and by extension the practice of design", and a growing number of IS researchers, such as Lee and Nickerson (2010) and Goldkuhl (2011), assert that pragmatism is the most appropriate theoretical paradigm to support DSR in particular. Overall, however, pragmatism remains more or less a niche perspective for grounding academic discourse and practice, often misunderstood despite its promise and potential (Brown, 2012; Godfrey-Smith, 2002; Ulrich, 2016).

Ulrich (2016, p. 5) argued that an important reason for this situation is that although pragmatism represents an important step forward in our understanding of (scientific) practice and has the potential to make a methodological difference that matters, its proponents have not been particularly successful in advancing methodological principles and corresponding conceptual frameworks that can be readily translated into improved practice. For example, while Dewey (1938) eloquently detailed how inquiry can be understood as a rigorous discursive process in which both contextual aspects of a situation and general standards for intelligent action (e.g., warranted assertibility as the end of inquiry) are productively intertwined to unfold the implications of ones assumptions and ideas about a situation in focus, it remains somewhat unclear how this understanding can be leveraged to support responsible practice. In a sense, one could argue that Dewey (1938) told us a lot about *what* to do, but less about *how* to do it.

As argued in the introduction section of this chapter, this lack of understanding regarding the implementation of responsible (scientific) practice is becoming increasingly pernicious in the face of growing technological power and the corresponding demand for more RI. Today, we recognize that basically all challenges of societal import should be viewed as *wicked problems*[9]—open-ended social problems that are ill-defined and "rely upon elusive political judgment for resolution" (Rittel & Webber, 1973, p. 160). This understanding extends the pragmatist insight that problems are not given, but constructed in relation to possible solutions as part of inquiry, with the recognition that, in the absence of a universally accepted theory of value, there is no one way of framing a wicked problem that is better than all the others, because "what satisfies one may be abhorrent to another, that what comprises problem-solution for one is problem-generation for another" (Rittel & Webber, 1973, p. 169). In this sense, "social problems are never solved. At best they are only re-solved—over and over again" (Rittel & Webber, 1973, p. 160). Thus, in practice, we face not only an *epistemological crux* associated with reaching the end of inquiry (i.e., how do we establish the warranted assertibility of our judgments in an increasingly complex and interconnected world?), but also an *axiological crux* associated with justifying the direction of inquiry itself (i.e., how do we justify our framing of a problem in a pluralistic world where diverse values are held?).

Reflecting on the life's work of C. West Churchman, Ulrich (1994) explored the implications of these two cruxes for the possibility of what he called *future-responsive management*, or what we might more generally call *responsible inquiry*, understood as the practice of securing "improvement that persists" (Ulrich, 1994, p. 29). As he observed, the epistemological and axiological cruxes pose fundamental cognitive and methodological difficulties for responsible inquiry that must somehow be dealt with in practice.

Regarding the epistemological crux, he used one of Churchman's classic examples, the inventory problem, to illustrate how it pushes us toward a *whole-systems* perspective:

> As Churchman demonstrates, it is quite impossible to design an optimal inventory policy for a manufacturing plant without considering all conceivable alternatives to holding an inventory. The best of the foregone opportunities represents an inventory policy's opportunity costs. How can we reasonably judge the degree to which an inventory policy is sound without knowing its opportunity costs? But any estimation of these opportunity costs bursts the definition of the inventory problem and requires us to investigate the larger system of the firm's opportunities. As the relatively simple inventory problem shows, we cannot rationally design improvement without assuming some theory about the nature of the total relevant system. (Ulrich, 1994, p. 27)

This logical requirement for *comprehensiveness* in framing and planning, and the cognitive demands associated with it, will only increase as our world becomes more interconnected and interdependent in the wake of long-term trends such as globalization and digital transformation. As we enter the Anthropocene (Lewis & Maslin, 2015), we must accept that even our local actions (e.g., driving a fossil fuel car) can and do have global and lasting impacts (e.g., contributing to climate change). Thus, if we are to recognize the responsibility that this situation demands of us (e.g., Jonas, 1984), we are increasingly forced to consider the entire Earth system as our reference system of concern

[9] "The adjective 'wicked' is supposed to describe the mischievous and even evil quality of these problems, where proposed 'solutions' often turn out to be worse than the symptoms" (Churchman, 1967, B-141).

(D. S. Wilson et al., 2023, p. 7). However, this puts us in a difficult epistemic position because we cannot reliably forecast the effects of our actions on the Earth system (e.g., Rockström et al., 2009), let alone on our societies and organizations (e.g., Consilience Papers, 2022; Hillis, 2010; Perrow, 2011), or even on ourselves as individuals (e.g., Tarafdar et al., 2013).

As Ulrich (1994) pointed out, in such a complex, ever-changing environment, we cannot solve the epistemological crux once and for all, because there may always be relevant considerations that we have simply overlooked. There is no comprehensively rational solution. Instead, our only hope is to engage in critique and critical discourse to *re*-solve it again and again. Specifically, he argued that we should find practical ways to foster and sustain a critical discourse that moves beyond a reductionist focus on forecasting the effects of actions, and expands its scope to include developing a better understanding of the situations in the first place and engaging with emancipatory critiques of our ways of responding to them. In this way, we can iteratively move toward an at least critically tenable resolution to a crux for which there is simply no comprehensively rational solution.

Looking at the axiological crux, Ulrich (1994) identified a very related problem. Given the inevitability of conflicting values in practice, all our decisions and actions have normative content that should be both ethically reflected and democratically legitimized. Improvement rarely, if ever, means improvement for all concerned and affected. For better or worse, judgments must be made about whose concerns are considered and how costs are distributed; sides must always be taken. As Ulrich (1994, pp. 28–29) observed, in a world of conflicting values, this means that "ethically defensible agreement on the standard of improvement" can quickly become the scarcest resource to manage.

For example, it is easy to see how the common practice of focusing scientific inquiry on developing solutions to general problems, largely in isolation from considerations of the needs of the wider world, leads to a convenient but ethically highly controversial situation: "Who cares about planets and their motions in a world of starvation? Instead of landing a man on the moon at a cost of several billions of dollars, why don't we feed some starving babies, or improve the education of our kids, or help preserve the natural environments?" (Churchman, 1995, p. 267).

In today's interconnected, conflicted, and rapidly evolving world, such complex value conflicts cannot be adequately addressed by traditional perspectives on ethics focused mostly on individuals and the moral quality of their individual actions (Jonas, 1984; Ulrich, 1994). Even our individual actions are connected to, and thus should also be viewed in light of, our global situation, which requires significant cognitive effort to unravel and systematic collective engagement to improve (e.g., Consilience Papers, 2022; Hillis, 2010; Levin et al., 2012; Ord, 2020). Similar to the epistemological crux, we therefore need to expand our concern beyond the immediate situation and consider actions also in relation to "the harm or improvement caused in the whole system" (Ulrich, 1994, p. 33). Thus, we are once again pushed to consider the entire Earth system as the appropriate reference system for many, if not most, of our ethical considerations.

In this way, the axiological and epistemological cruxes of wicked problems are inherently intertwined, both calling for a whole-systems perspective. However, as noted above, such a perspective can never produce a comprehensively rational solution, since there may always be relevant considerations that we have simply overlooked, whether they be relevant facts (cf. epistemological crux) or values (cf. axiological crux). Thus, for Ulrich (1994, p. 35), "the implication of the systems idea is not that we must understand the whole system but rather that we need to deal critically with the fact that we never do". This is the key insight underlying his later calls for a "critical turn" in pragmatism (e.g., Ulrich, 2006, 2016), the idea that responsible inquiry requires the open and self-reflective handling of the inevitable limitations of our claims.

To further explicate this vision for a critical pragmatism, Ulrich (2006) developed a critical pragmatist perspective on professional ethics. For him, critical pragmatism views the task of responsible practice as one of finding practical ways to balance the demands of both *ethical universalism*[10] and *ethical contextualism*[11]: "If we want to support ethical practice, there is no way round striking a methodological balance between these two poles" (Ulrich, 2006, p. 73).

Ulrich's (2006) proposal for achieving this task was that ethical universalism and ethical contextualism both have critical potential that can be pitted against each other in order to work toward critically tenable resolutions that balance both concerns. This reflects the recognition that responsible inquiry is not only inherently situated (i.e., the contextualist pole) but also, as the preceding discussion illustrates, ideal-seeking (i.e., the universalist pole), and it suggests that the balancing between these poles should be implemented "[...] in such a way that each can play a critical role in dealing with the unavailability of a positive solution to the other" (Ulrich, 2006, p. 73).

As Ulrich (2006) argued, the implementation of this proposal can be supported through his earlier work on *Critical Systems Heuristics*[12] (CSH; Ulrich & Reynolds, 2010) and its core principle of *boundary critique*. The key insight behind CSH, and boundary critique in particular, is that the justification of *boundary judgments* (i.e., how we draw boundaries around a situation) is the core methodological problem that reflective and responsible practice must address. Boundary judgments are the product and source of our *selectivity* with respect to the observations and evaluations we make, and thus must be subjected to ethical scrutiny (Ulrich, 2006, pp. 74–77): From the perspective of ethical contextualism, the main concern with selectivity is the arbitrariness of context (cf. epistemological crux); from the perspective of ethical universalism, the main concern is partiality (cf. axiological crux).[13]

Fortunately, however, the concerns of ethical contextualism and ethical universalism can be used to keep each other in check. Arbitrariness of context can be kept in check by reflecting on the observations that are being made and pushing out to consider facts pertaining to the whole system (cf. ethical universalism). Partiality can be kept in check by reflecting on the evaluations that drive problem framing and zooming in on the particular values that shape a situation (cf. ethical contextualism). This insight is used in the practice of boundary critique, where these two concerns are brought together to support a systematic process of surfacing and unfolding of selectivity in claims with respect to the "facts" (i.e., empirical circumstances) and "values" (i.e., normative considerations) on which they are based (Ulrich, 2006, p. 76).

As visualized in Figure 3, the practice of boundary critique can be made more intuitive by understanding it as a process of working through the "Eternal Triangle" of boundary judgments, observations, and evaluations (Ulrich, 2006, p. 77).

[10]Ethical universalism refers to the idea that action is ethical if its maxim can be made the principle of a universal law (cf. categorical imperative). In practice, it calls for the establishment of basic ethical principles that apply to all situations.

[11]Ethical contextualism refers to the recognition that practice is contextual, and the assertion that the ethicality of actions must therefore be evaluated in light of the specifics of a situation. In practice, it calls for the context-sensitive application and evaluation of ethical principles.

[12]CSH is a framework for reflective professional practice based on a set of 12 questions that can help to make explicit the everyday judgments we rely on to understand situations and identify improvements (Ulrich & Reynolds, 2010, pp. 243–244).

[13]In this sense, boundary judgments can be argued to be a unifying concern behind the epistemological and axiological cruxes, both of which boil down to the question of how we can justify the selectivity of our inquiries with respect to the facts we observe on the one hand (i.e., the epistemological crux), and the values we use to evaluate on the other (i.e., the axiological crux).

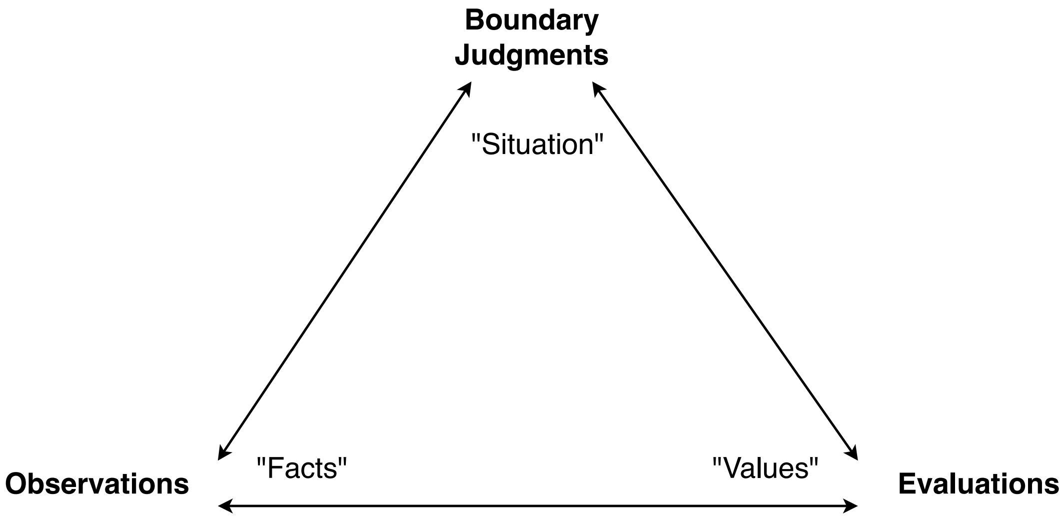

Figure 3: Boundary Critique as the Process of Working Through the"Eternal Triangle" of Boundary Judgments, Observations, and Evaluations (based on Figure 2 from Ulrich, 2006, p. 77).

To better understand the Eternal Triangle, let us consider a practical scenario where a team is tasked with developing a new AI system for customer service. Initially, the scope of the project was set to prioritize efficiency and cost reduction as key values, resulting in a focus on minimizing the need for human customer service agents. Working within this scope, the team identified facts related to the initial development costs and expected time savings. However, while testing a prototype, new observations came to light. They discovered that the AI system's responses lacked empathy and negatively impacted customer satisfaction. This new observation prompted a shift in the teams' boundary judgments. They recognized the need to adjust the scope of the project to include a more empathetic customer experience as a requirement for the AI system. As a result, the reference system within which they viewed the AI system changed to include both efficiency and empathetic customer interactions. This, in turn, led to an adjustment in their values. Whereas efficiency and cost reduction had been their primary values, they now wanted to place greater emphasis on customer satisfaction and empathy. This change in values also affected how the team assessed the facts. For example, the initial development costs were now seen to be less important when compared against the potential loss of customers due to dissatisfaction with the AI system.

This example illustrates how changes in one element of the Eternal Triangle trigger adjustments in the other two, leading to an iterative and interdependent process of boundary critique in practice. Moreover, it also illustrates how some forms of boundary critique happen naturally all the time. However, rather than seeing boundary critique as simply a descriptive process, Ulrich's (2006, pp. 77–79) key insight was that it can be supported and systematically used for ethical and emancipatory purposes, and thus help to facilitate the development of at least critically tenable resolutions to problematic situations.

In particular, *ethical boundary critique* can be used to broaden the scope of considerations by surfacing and unfolding the ethical implications that particular boundary judgments have for all conceivable parties affected by a situation, rather than just the intended beneficiaries and parties directly involved. This "looking beyond" can raise ethical awareness of the inevitable selectivity of problem framings and improvement proposals, and thus help to ensure that "the concerns of those on the 'losing side' of an ethical position are given due consideration" (Ulrich, 2006, p. 78).

In addition, *emancipatory boundary critique* can be used to empower all affected parties to voice their concerns in a cogent and competent manner. For this to happen, Ulrich (2006, p. 79) argued, it is enough for affected parties to become aware of the inherent selectivity of boundary judgments, so that they can begin to question the particular boundary judgments being made in a situation, and to demonstrate that there may be other options. Once the illusion of objectivity is shattered, "we discover boundary judgments at war" (Ulrich, 2006, p. 79). In this way, emancipatory boundary critique "can secure to all the parties a gain in *symmetry of critical competence*" (Ulrich, 2006, p. 79).

Taken together, ethical and emancipatory boundary critique can help to open up the necessary spaces in which ethical reflection and critical discourse can unfold, no more and no less. In this sense, boundary critique is certainly not a panacea for securing ethical and responsible practice. It does not offer general answers, but only points to a lack of them. It does not produce convincing justifications, but only aims to facilitate sufficient critique. Moreover, as a practical tool for creating conditions conducive to ethical reflection and critical discourse, it is necessarily incomplete in addressing the challenges that responsible practice must overcome. As such, it can only be a starting point and not an end point for responsible practice.

In conclusion, critical pragmatism responds to the recognition of the complex interplay of epistemological and axiological dilemmas in addressing wicked problems by calling for a democratic vision of society in which all people are empowered to resolve their conflicts and improve their conditions in peaceful ways through critical discourse (Ulrich, 2006, p. 80). To this end, it proposes systematic boundary critique as a practical tool that can help to foster ethical awareness, promote dialogue, and navigate the complex terrain of responsible inquiry. In this sense, critical pragmatism does not offer easy solutions, but rather underscores the ongoing, iterative nature of ethical reflection and responsible practice in our interconnected world. As we grapple with increasingly complex and global problems, it reminds us of the value of open dialogue, ethical reflection, and self-awareness in the ongoing quest for at least critically tenable resolutions to the challenges we face.[14] In the next section, I discuss why it can serve as an appropriate theoretical foundation for this dissertation.

2.4.3 Critical Pragmatism as a Theoretical Foundation for this Dissertation

Pragmatism has long been recognized as the philosophy best suited to ground DSR (e.g., Goldkuhl, 2011; Hovorka, 2009; Lee & Nickerson, 2010; Sjöström, 2010). Its inherent focus on practical consequences and problem solving (Dewey, 1938) aligns perfectly with the goals of DSR, which aims to build and evaluate innovative artifacts that help resolve relevant real-world problems (Hevner et al., 2004). However, DSR has not yet fully exploited this synergy. None of the foundational works on DSR in the IS community has provided "systematic and clear articulations of the philosophical foundation of DSR"

[14] As a point of clarification, I want to emphasize that critical pragmatism remains agnostic about what a critically tenable resolution might look like. For example, it may be the case that after extended engagement with a problem situation, it turns out that what was thought to be a problematic situation is actually not so problematic and does not warrant the investment of additional resources in the matter, given the opportunity costs of doing so. This insight seems particularly relevant to the field of IS research, where the practical and societal relevance of much of the work done continues to be critically discussed (e.g., Herwix & Haj-Bolouri, 2021; Moeini et al., 2019; Ram & Goes, 2021).

(Deng & Ji, 2018, p. 6). As discussed in the previous section on the historical development of DSR, this has contributed to a state of confusion about the possibilities and limitations of DSR, ultimately obscuring its full potential.

In this dissertation, I draw on critical pragmatism to clear up the confusion and fill this gap. I have chosen critical pragmatism rather than classical pragmatism for this task because it not only articulates the challenges of responsible practice more explicitly, but also offers practical, methodological suggestions for dealing with them. Given the focus of this dissertation on the development of a responsible DSR ecosystem, this seemed to be the distinguishing factor in its favor. However, critical pragmatism would itself point to the inherent selectivity of this choice. It may well be possible to find additional considerations that would tip the balance in favor of another philosophical perspective. In this sense, this choice cannot be justified in a fully rational way. Rather, I must ask the reader to judge its merits on the basis of its consequences for the material presented in this dissertation.

Research Approach

3.1 Overview

This dissertation applied a critical pragmatist perspective. This placed an emphasis on recognizing the inherent limitations and fallibility, and thus the iterative and provisional nature of this work. Furthermore, given the goal of initiating and supporting the development of a responsible DSR ecosystem, this work was inherently oriented toward collaborative and open engagement with the IS research community. I tried giving recognition and space to the diversity of voices in the field and following Open Science practices (e.g., Doyle & Luczak-Roesch, 2019) where appropriate. I have also tried to follow the practice of "Eating Your Own Dog Food" (2023) by applying the concepts and recommendations I have developed in the course of my own work.

Figure 4 reconstructs the overarching methodological structure of this dissertation in reference to the *Multi-Layered Development Model for Responsible DSR* (MLDM), which will be introduced in detail in Publication 2 of this dissertation.[1] The overall purpose of this dissertation was addressed through three interrelated inquiries, with the first two taking an overt meta-research perspective and the third taking an applied responsible DSR perspective. The order of the inquiries implies a logical progression towards the resolution of the overarching problem situation:

1. The first inquiry institutes the initial framing of the problem (i.e., a proof-of-problem) based on a panel discussion substantiated with a broad, narrative literature review and thereby initiates the problem stage of this dissertation.

2. The second inquiry updates the problem's framing (i.e., proof-of-problem) and develops concepts that suggest first steps towards resolution (i.e., a proof-of-concept). This is achieved through multi-grounded theory development based on 41 interviews with stakeholders from the IS research community. This inquiry, thus, marks the transition from the problem stage to the concept stage in this dissertation.

3. The third and final inquiry starts to put the developed concepts to the test by deploying them in an applied responsible DSR discourse. This is done by instituting a real-world problem (i.e., proof-of-problem) based on a scenario development exercise and simulation experiments. This inquiry substantiates the concept stage of this dissertation by initiating the problem stage of an applied responsible DSR discourse.

[1] In the name of transparency, I want to emphasize that in the following I will present an idealized reconstruction of the actual research process. In practice, all of the inquiries were conducted more or less in parallel, with considerable exchange between them. For example, I could also call Inquiry 2 the first inquiry because it started before Inquiry 1 and actually helped to initiate it. However, I chose the labels the way I did because Inquiry 1 was completed much earlier than Inquiry 2. Similarly, Inquiry 3 unfolded for some time in parallel with Inquiry 2. Despite these inaccuracies, the idealized reconstruction in Figure 4 seems useful for conveying the intended logical relationship between the inquiries that form this dissertation.

Altogether, this methodological reconstruction reflects my commitment to a critical pragmatist perspective. It allows me to self-critically examine and communicate how the findings presented in this dissertation developed iteratively over a series of inquiries that build on each other. In the following, I will summarize each of the inquiries' research approaches in more detail.

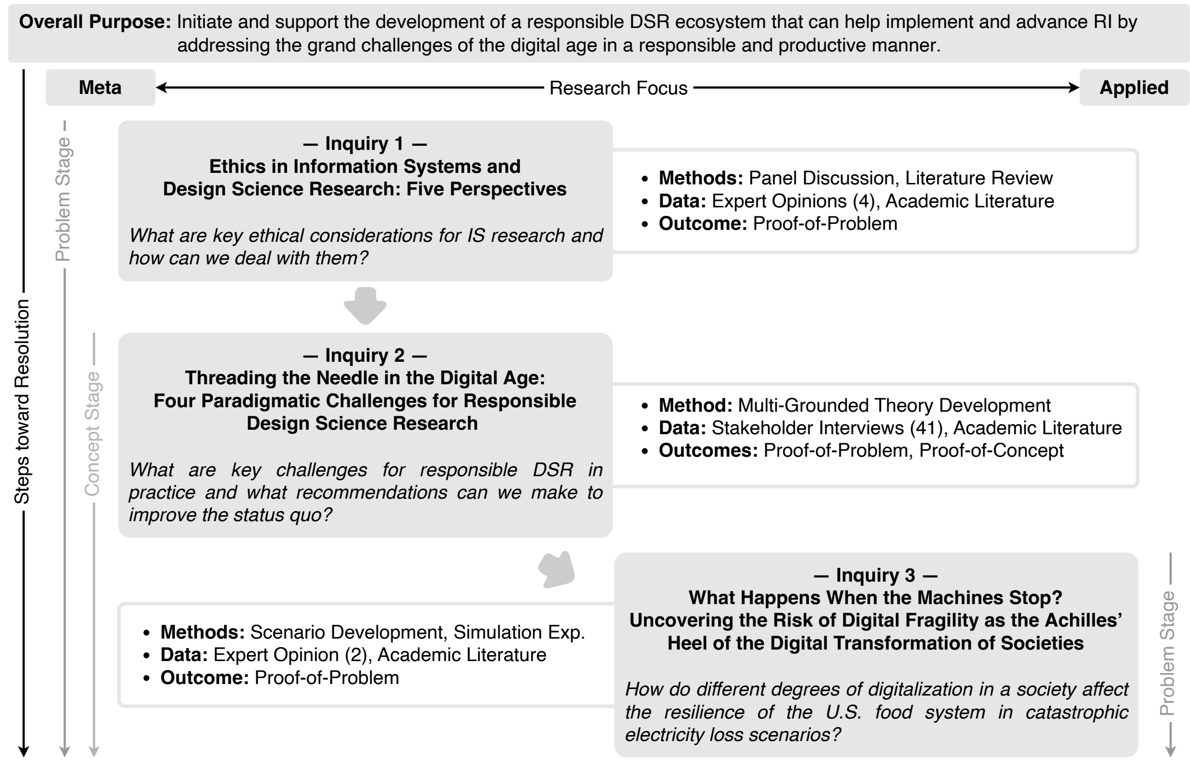

Figure 4: Overview of the Methodological Structure of this Dissertation.

3.2 Inquiry 1: Panel Discussion and Literature Review

The first inquiry instituted an initial problem framing for this dissertation by investigating the research question: *What are key ethical consideration for IS research and how can we deal with them?* We were motivated to focus on this question due to an increasing recognition of ethics as a key issue of our times. As the reach of digital technologies and systems expands and the impact of digital transformation on our societies grows, we are under increasing pressure to have our work shaped not only by practical or economical but also by ethical considerations as well (e.g., Maedche, 2017).

We addressed the research question by combining reflections from a panel discussion on "Ethics in DSR" at the *International Conference on Design Science Research in Information Systems and Technology* (DESRIST 2020), featuring four highly experienced IS scholars, with insights from a broad, multidisciplinary, narrative literature review (Boell & Cecez-Kecmanovic, 2014a, 2014b). This approach allowed us to develop a novel, multifaceted, yet integrative account of five interrelated perspectives (i.e., philosophy, design, science, IS practice, and panelist perspectives) on ethical considerations for IS research that can help us better understand the opportunities and challenges we face with respect to ethical IS research practice.

Methodologically, we followed Wessel et al.'s (2021) key recommendations for organizing panels and writing panel reports, which emphasized the need to embed the perspectives presented at the panel into the larger context of the relevant academic discourse. We achieved this integration by conducting a broad, multidisciplinary, narrative literature review informed by Boell and Cecez-Kecmanovic's (2014a) hermeneutic framework for literature reviews. This framework positions the development of understanding as the key concern that literature reviews must address and emphasizes that this is best achieved through iterative, hermeneutical processes of engagement with the literature. In particular, Boell and Cecez-Kecmanovic (2014a) advocate for a deliberate and measured literature searches focused on the identification of key literature, as opposed to mindless systematic searches of literature databases that produce unmanageable amounts of mostly meaningless results. A key insight we gained from their work was their suggestion that broad, multidisciplinary reviews should focus on relevant, high-quality secondary literature, such as authoritative encyclopedia articles. This helped us to ensure that we were getting good coverage of key considerations, rather than getting lost in the weeds of highly complex research areas that were at the outset unfamiliar to us.

Overall, this inquiry has produced a first version of, what I call, a *proof-of-problem*. A proof-of-problem refers to the establishment of a problem framing that has been shown to be reasonably defensible in critical discourse, for example, through a peer-review process. As I will discuss in Publication 2 of this dissertation when I introduce the MLDM, I propose that attaining a proof-of-problem should be an important first step in responsible DSR, because our understanding of the problem ultimately determines the results we can achieve (Rai, 2017b). Therefore, not only at the beginning, but also throughout a responsible DSR initiative, it is crucial to subject our problem framing to sufficient criticism. In this sense, the open-ended, discursive process of publishing Inquiry 1 in the *Communications of the AIS* helped me to develop, challenge, and justify the initial problem framing of this dissertation in relation to the expectations and interests of the IS research community.

3.3 Inquiry 2: Multi-Grounded Theory Development

The second inquiry continued the exploration of the relationship between ethics and IS research practice begun in the first inquiry. This time the focus was on developing a deeper understanding of what the notion of *responsible DSR* should entail. It was motivated by the recognition from the first inquiry that more work was needed to create conditions more conducive to ethical and responsible research practices before the IS research community could realize its full potential in this regard. In order to help improve this situation, I investigated the following research question: *What are the key challenges for responsible DSR in practice and what recommendations can we make to improve the status quo?*

To address this research question, I developed a multi-grounded theory (Goldkuhl & Cronholm, 2010, 2018) of four key paradigmatic challenges for responsible DSR in practice, as well as recommendations for improving the status quo based on this newfound understanding. To ground this effort, I conducted 41 in-depth interviews with relevant parties from the IS community and drew on academic literature from a variety of fields, including management, cognitive science, political economy, innovation studies, and more.

Methodologically, the primary goal was to achieve Goldkuhl and Cronholm's (2010) ideal of multi-grounding, that is, alignment between the empirical data and the emergent theory (i.e., empirical grounding), alignment between the emergent theory and external theories (i.e., theoretical grounding), and alignment between concepts within the emergent theory itself (i.e., internal grounding). Therefore, I followed Goldkuhl and Cronholm's (2010) recommendation to supplement the theory generation process with explicit grounding processes to ensure the multi-grounding of the emerging theory. Specifically, I engaged in an iterative and flexible theory generation process that included *inductive coding*, *conceptual refinement*, *pattern coding*, and *theory condensation* steps, consistent with the pragmatic strand of the grounded theory research tradition (Bryant, 2017). This was supplemented by explicit grounding practices, including moving to a second research iteration when I found that the research interests had evolved based on a growing understanding of the situation, constantly relating the coding concepts not only to the interview data but also to external theories, and transparently documenting the prevalence of the paradigmatic challenges in the interview data to test the internal logic of the emergent theory and ensure its fit with the data.

Overall, this inquiry advanced our understanding of the challenges associated with the practice of responsible DSR and helped me take the first steps toward developing promising strategies to address them. Using the terminology of the MLDM, this inquiry helped me to refine the proof-of-problem to the point where I felt comfortable moving to the concept stage. This meant starting to work on possible solutions with the goal of generating working *proof-of-concepts* (i.e., prototypical solutions). To this end, four specific solution proposals, such as supporting the iterative development of responsible DSR according to the MLDM, were developed as starting points for further experimentation in the future. Thus, while the first steps have been taken to develop proof-of-concepts, their status must be considered tentative, as further experimentation would be required to determine whether they can successfully address the requirements of the problem situation.

3.4 Inquiry 3: Scenario Development and Simulation Experiments

The third inquiry shifted the focus from a meta-research perspective on responsible DSR to applying and testing the insights gained so far by addressing an important real-world problem situation outside of academia. This required a compelling proof-of-problem to initiate a prototypical responsible DSR discourse. In line with the core interest of this dissertation in RI, we decided to explore the ambivalent role of the ongoing digital transformation for the resilience of our societies. While most research on digital transformation emphasizes its potential to contribute to the generation of economic benefits (e.g., Vial, 2019) and, increasingly, its potential to contribute to societal resilience (e.g., Boh et al., 2023), there has also been a reckoning of sorts, with large-scale disruptions and critical incidents involving digital systems occurring on an almost regular basis (e.g., Perrow, 2011; Sanger & Perlroth, 2021; Seger et al., 2020; Wolff, 2020) and concerns about existential risks from AI increasing in intensity in recent years (e.g., Center for AI Safety, 2023; Ord, 2020). Against this backdrop, we decided to problematize the potential dark side of the ongoing digital transformation of our societies by investigating the research question: *How do different degrees of digitalization in a society affect the resilience of the U.S. food system in catastrophic electricity loss scenarios?*

We addressed this research question through a three-phase research process that combined scenario planning in the intuitive logic school (Amer et al., 2013; Wilkinson et al., 2013) to identify plausible *catastrophic electricity loss (CEL) scenarios*[2] (first phase) with the development of a customized system dynamics (SD)-based simulation (Fang et al., 2018) to examine the impact of such scenarios on the resilience of the U.S. food system given different degrees of digitalization in society (second phase). The process concluded

[2]With CEL scenarios we mean blackouts covering a large geographic area with the size of multiple US states or EU countries, affecting at least 90% of the population and lasting for at least 30 days.

with a final phase of synthesis and reflection (third phase). This approach allowed us to combine the flexibility of scenario planning with the methodical and audit-able nature of simulation-based research.

Methodologically, in the first phase, we used the scenario planning principles of the intuitive logic school (Amer et al., 2013) to develop plausible CEL scenarios. That is, we viewed the scenarios as tools for understanding complex, rare, and uncertain events rather than as tools for predicting them. This was reasonable because, although CEL scenarios are inherently complex, rare, and uncertain events that are difficult to predict, there is still value in trying to better understand and prepare for them because being difficult to predict does not mean that their occurrence is unlikely (Goodwin & Wright, 2010). In this sense, our scenarios should be seen as fictions and stories that can be used to challenge assumptions, start discussions, and work toward shared frames of interpretation (Goodwin & Wright, 2010; Wilkinson et al., 2013). The development of these scenarios was informed by the academic literature, the expertise of the research team, and feedback from two external experts, and took place in three steps:

1. Development of a comprehensive scenario framework that provided a broad overview of the different types of catastrophic risks that could lead to CEL.

2. Selection of two scenarios that seemed plausible and relevant to our inquiry.

3. In-depth analysis of the causal logic underlying the selected scenarios.

The selected scenarios then served as critical inputs for the second phase, which involved adapting an existing, peer-reviewed SD-based simulation model of the U.S. food system (Huff et al., 2015) to specifically assess the resilience of the U.S. food system in the face of CEL scenarios. The decision to use an SD model was driven by its potential to rigorously explore complex systems, allowing for in-depth analysis of the consequences of different assumptions and choices. We used the guidelines and tests suggested by Fang et al. (2018) to ensure the validity and usefulness of our simulation model. We then conducted two simulation experiments, based on Monte Carlo simulations with thousands of iterations, to examine how different degrees of digitalization in society would affect the resilience of the U.S. food system in the selected CEL scenarios.

In the third and final phase, we reflected on the underlying logic of the selected scenarios and the dynamics we had observed in the simulation experiments. As a result of this process, we developed the concept of *digital fragility* as a more general explanation of how the digital transformation of our societies can increase systemic risks of catastrophic outcomes, and also identified a tentative set of promising strategies that could help to reduce the risk of digital fragility.

Overall, this inquiry has helped me test some of the ideas about responsible DSR by allowing them to guide us in developing a proof-of-problem that can serve as a starting point for a responsible DSR discourse about an important real-world challenge of the digital age. It is too early to tell how successful this experiment will be. However, the first steps taken already demonstrate at least the basic feasibility of responsible DSR as outlined in this dissertation. In particular, the MLDM has proven helpful in planning this third inquiry. Nevertheless, more work is needed to continue the efforts begun in this inquiry. Establishing a proof-of-problem is only the very first step toward resolution. Responsible DSR requires continued engagement with the problem situation until at least a critically tenable resolution is found.

Main Results of this Dissertation

4.1 Inquiry 1: Opportunities and Challenges for Ethics in Information Systems and Design Science Research

The first inquiry not only reviewed five perspectives on ethics (i.e., philosophy, design, science, IS practice, and panelist perspectives) that are relevant to IS research and practice, but also identified, based on panelist input, two key challenges for ethics in IS practice and three promising opportunities for addressing them. The two key challenges to ethics in IS practice were: (1) *a lack of skills and resources*, and (2) *value conflicts*. The three promising opportunities to support greater engagement with ethical considerations in IS research were: (3) *encouraging value reporting*, (4) *engaging in problem prioritization*, and (5) *improving the research infrastructure*. Each of them will be summarized in turn.

First, staying competitive in the digital age requires significant resources. This constant pressure can divert attention and resources away from ethical decision-making. This situation is increasingly untenable in times when technological risks and potential unintended consequences can have far-reaching impacts. A case in point is IBM's Watson, which was marketed as a revolution in cancer care but has been criticized for poor performance and an inability to deliver anything close to what was promised, wasting precious resources and potentially even endangering patients' lives in the process (Ross & Swetlitz, 2017; Strickland, 2019). Despite ethical codes such as the *ACM Code of Ethics and Professional Conduct* (2018), organizations often lack the necessary support for ethical decision-making. This lack of skills and resources for ethical concerns is a significant hurdle, calling for the development of effective tools and infrastructure to facilitate ethical and competent behavior.

Second, a major challenge arises from the diversity of value systems in modern societies, which often leads to tensions over conflicting values. For example, Awad et al.'s (2018) moral machine experiment illustrates how people from different cultural backgrounds have different moral intuitions in scenarios involving autonomous cars. This raises pivotal ethical questions about algorithm design. Should algorithms be tailored to specific cultural norms? Can autonomous car manufacturers freely choose their algorithms, or should there be a universally mandated standard? In addition, significant resources will be needed to verify that these algorithms are consistent with their intended goals. Navigating these ethically charged value conflicts will become increasingly important as we digitize and manage more of our world with algorithms. While existing research can offer some guidance, addressing these conflicts remains a challenging endeavor that demands expertise, deep engagement, and iterative refinement.

Third, *"value reporting"* was identified as a promising intervention to enhance ethical awareness in IS research. This involves explicitly stating the underlying value systems used or considered by the authors. This could be included as a dedicated paragraph in the theory development or methods section, or as a separate section or appendix. Value reporting promotes openness and transparency, and encourages authors to reflect on how their work aligns with their stated values. This increased transparency could help the IS community become more self-reflective, transparent about potential value conflicts, and sensitive to new research opportunities. For example, it could reveal which values IS researchers consider relevant to their work. Encouraging authors to reflect on their core values could also lead to innovative research perspectives and more ethically conscious problem framing. In addition, value reporting could help reviewers better understand authors' intentions, especially on controversial or ethically sensitive topics. As such, it could enable editors and reviewers to make more informed decisions and provide more targeted suggestions.

Fourth, it was emphasized how a more systematic, ethically conscious approach to selecting, framing, and prioritizing research problems could help increase the relevance of the IS field, since the relevance of our work depends on the problems it addresses. Ethics can guide us in evaluating research problems from different perspectives (Herwix & Haj-Bolouri, 2020, 2021), thus providing a sounder basis for prioritizing research topics in IS research and helping us to respond to calls for a more strategic focus in IS research (Gable, 2020; Winter & Butler, 2011). Problem prioritization could be implemented by individuals and research teams when selecting the problems they want to focus on in their work, or more institutionally, by top journals offering dedicated sections and special issues for critical discussion of evolving research priorities.

Fifth, building on the previous findings, a final key opportunity for IS research was identified: building a comprehensive research ecosystem that promotes ethical behavior and productive work. As suggested by Winter and Butler (2011), the IS community, with its expertise, could play a central role in designing and building the necessary infrastructure for such a research ecosystem. For example, by developing strategies, prototypes, and design principles that help address societal challenges in a responsible and productive manner. The potential impact of such an effort could be substantial, as improvements in the IS research ecosystem could also benefit other domains and help shape academic standards of success. In particular, by improving our ability to conduct ethical research effectively, we can create positive feedback loops that enhance the quality and legitimacy of our work. For example, other research fields, such as astronomy, have already demonstrated how sociotechnical research infrastructure can significantly increase research quality and productivity (Watson & Floridi, 2018); innovative educational formats can develop ethical awareness and competence (Berti et al., 2021; Parks-Leduc et al., 2021); and institutional support for preprints, including anonymous but open peer review, can improve research practices and outcomes (Fu & Hughey, 2019; Ross-Hellauer et al., 2017). Thus, the IS community has much to gain and little to lose by proactively developing and improving its infrastructure.

Overall, the key findings of the first inquiry point to the importance of a more deliberate integration of ethical considerations into the IS research enterprise. In particular, the development of a research ecosystem conducive to ethical behavior and practice was identified as a major opportunity, but also a key challenge, for IS research in the digital age. This finding motivates and prefigures the second inquiry, which looks more closely at the requirements that such a research ecosystem should address and, based on this, makes recommendations for its implementation.

4.2 Inquiry 2: Paradigmatic Framework for Responsible Design Science Research

The second inquiry took the first steps toward actively initiating the development of a responsible research ecosystem by laying a conceptual foundation for responsible DSR. In particular, it developed a comprehensive paradigmatic framework that conceptualizes responsible DSR as pragmatic inquiry that must address four interrelated *paradigmatic challenges*:

- *Ontological challenge*: Making warranted boundary judgments,
- *Epistemological challenge*: Using, producing, and growing a body of useful knowledge,
- *Axiological challenge*: Managing conflicting values and positions, and
- *Methodological challenge*: Developing, selecting, and integrating a portfolio of practices.

These challenges are *paradigmatic* in nature because they explicitly address the four major dimensions along which research paradigms tend to be characterized (e.g., Chen & Hirschheim, 2004; Deng & Ji, 2018; Iivari, 1991; Iivari, 2007; Iivari et al., 1998; Lincoln et al., 2017; Morgan, 2007): *Ontology* (what are the nature and relations of being?), *Epistemology* (what are the nature and grounds of knowledge?), *Axiology* (what are the nature, types, and criteria of values and of value judgments?), and *Methodology* (what are the nature and body of methods in this particular field?). However, rather than characterizing responsible DSR in terms of specific answers to the questions posed by these dimensions, as has been done in previous paradigmatic analyses, this framework stands apart by focusing on the nature of the questions themselves: To what key challenges in responsible DSR practice do the ontological, epistemological, axiological, and methodological dimensions refer? This has been useful and appropriate because it allowed me to propose a framework that promotes *consilience* (E. O. Wilson, 1999) by providing an integrative perspective on DSR which is *flexible* enough so that it is applicable to the broad diversity of DSR (Rai, 2017a) as well as *open*, accessible, and transparent to stakeholders outside of the IS field. It is also *actionable* and *comprehensive* in addressing key challenges of DSR practice. Yet, the framework remains *parsimonious* by focusing only on the key challenges that are most fundamental to *responsible* DSR practice.

The paradigmatic challenges are *interrelated* in that each of them depends on crucial information produced by engagement with the others. For example, in Publication 2 I use practical examples to illustrate how *making warranted boundary judgments* (the ontological challenge) depends on access to relevant *knowledge* (produced as a result of engaging with the epistemological challenge), *values* (negotiated as a result of engaging with the axiological challenge), and *practices* (selected as a result of engaging with the methodological challenge), while all the other paradigmatic challenges in turn depend on useful *framings* established by warranted boundary judgments.

Taken together, the paradigmatic challenges represent the heart and mind of responsible DSR inquiry, which seeks to transform problematic situations that are *indeterminate* and *beyond our current capacity* to deal with into resolved ones (Dewey, 1938). In practice, this means engaging with wicked problems that are to some degree *contested, complex*, and *uncertain* (Rittel & Webber, 1973; Wanzenböck et al., 2020). Such situations are *resolved* to the extent that inquiry has transformed them so that they are *sufficiently defined* and *within our existing capacities* to be dealt with as a whole (Dewey, 1938).

This means that all relevant stakeholders deem the situations to have become sufficiently *acceptable* rather than contested, *understandable* rather than complex, and *predictable* rather than uncertain (Wanzenböck et al., 2020). Figure 5 provides a visual summary of the paradigmatic framework for responsible DSR. Publication 2 offers a more in-depth look at the framework including detailed discussions of each of the four paradigmatic challenges.

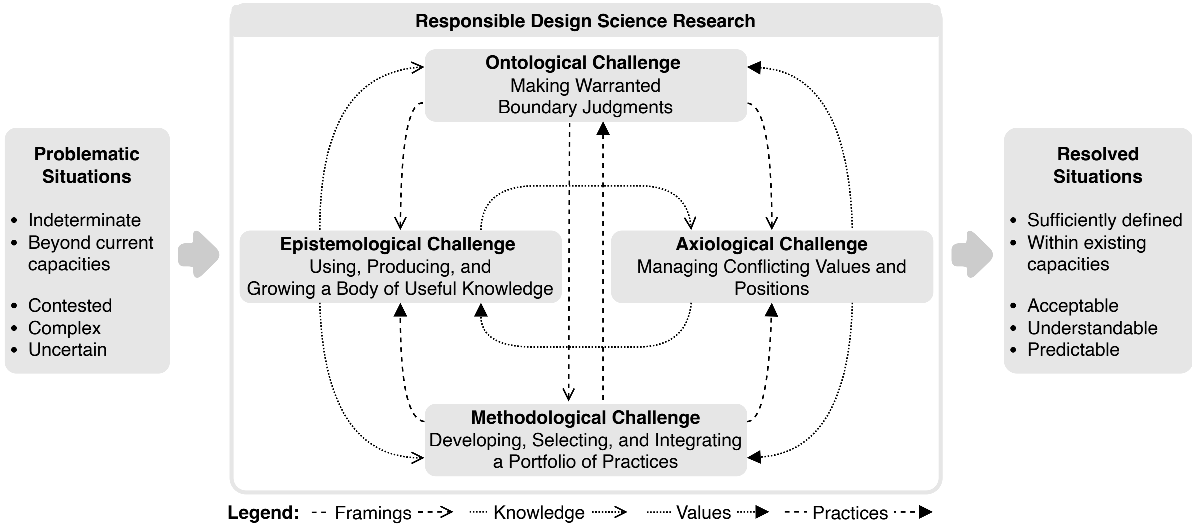

Figure 5: The Paradigmatic Framework for Responsible DSR.

In addition to the paradigmatic framework itself, the second inquiry also produced a set of four practice-oriented recommendations that outline more practical steps that can be taken to help bring about a responsible DSR ecosystem. In particular, the recommendations were designed to address the four paradigmatic challenges of responsible DSR in an integrated manner. They have been informed by a deep engagement with the interviews, the academic literature, and the framework for responsible DSR. As such, they provide an informed starting point for cautious experimentation and further critical discussion; they are not intended to be a panacea. Further work will be needed to test and refine the recommendations in practice. An overview of the four recommendations is provided in Table 2. A more detailed discussion is presented in Publication 2.

Overall, the key findings of the second inquiry articulate an inspiring vision for a responsible DSR ecosystem and a set of plausible starting points for its realization. However, significant implementation and testing of the proposals was beyond the scope of this inquiry. In keeping with the collaborative, open, and iterative spirit of responsible DSR and the MLDM in particular, the vision and proposals are opened up to critique and community discourse before taking bold steps toward implementation. Involving 41 members of the IS community in the development of the initial proposals was an important first step, but further implementation and testing should be informed by even broader forms of community involvement and participation. In this spirit, the third inquiry begins to establish a responsible DSR exemplar by developing a proof-of-problem for a societally relevant challenge of the digital age and seeking to initiate a collaborative, open, and iterative responsible DSR discourse around its resolution.

Table 2: Recommendations for Evolving a Responsible DSR Ecosystem.

Recommendation	**Description**
Establish Systematic Boundary Discourse	Systematic boundary discourse refers to the systematic use of boundary critique to guide us and our stakeholders in reflecting on and consciously (re)constructing our boundary judgments in responsible DSR practice. It can help us to become more responsible in the identification and selection of our research problems, as well as in the development and evaluation of our research projects.
Support Responsible Multi-Layered Development	Responsible multi-layered development is a high-level DSR process that views responsible DSR as a collective activity involving an entire research ecosystem. It identifies four layers of research that build on top of each other as responsible DSR moves towards resolving problematic situations. It aims for a more responsible coordination between relevant parties and a more productive allocation of resources.
Facilitate Polycentric, Multi-level Cooperation and Value Co-creation	How to effectively facilitate polycentric, multi-level cooperation and value co-creation at scale is not only one of the most pressing societal issues of our time, but also a question we must find an answer to if we are to develop a sustainable, responsible DSR ecosystem. I suggest that learning-by-doing in the form of the development and ongoing maintenance of a responsible DSR ecosystem is an appropriate way forward. Several existing insights and frameworks can be drawn upon to support this endeavor.
Develop Pragmatism, Reflexivity, Sensitivity, and Anticipation	Responsible DSR is a program that requires a strong methodological portfolio and the necessary skills and understanding to use it productively. Pragmatism provides a proven theoretical foundation and motivation for these efforts. Reflexivity opens up new perspectives that support its progress. Methodological and theoretical sensitivity are key competencies it requires. Anticipatory thinking and responsive action are what it seeks to develop.

4.3 Inquiry 3: The Risk of Digital Fragility

The third inquiry directly addressed a key challenge of the digital age: *What happens when the machines stop?* As the ongoing digital transformation of our societies pushes us to give machines an increasingly central role in our societies, we become more and more dependent on these machines to function within the parameters of their design. But what happens when they cannot? What happens when a major power outage (i.e., CEL) takes a large portion of the machines we increasingly depend on offline? Through scenario development and simulation experiments, we explored this question to understand the impact of a society's degree of digitalization on its resilience (particularly the resilience of its food system) in CEL scenarios.

The results provide evidence for the hypothesis that the degree of digitalization in a society does indeed contribute to the severity of the impact of CEL scenarios. Thus, our work shows how the ongoing digital transformation of our societies, while potentially beneficial, also seems to make us more vulnerable, at least in CEL scenarios. Building on these results, we have developed a more generalized account of this situation, visualized in Figure 6.

Our generalized conceptualization identifies the *standardization* and *interconnectedness* of digital systems and services as the key characteristics that give rise to the observed dynamics. While the standardization of digital systems and services allows for efficient implementation and updating of functions, the interconnectedness of digital systems and services via the Internet allows for efficient, fast, and long-distance transmission of information. Taken together, these characteristics drive complex dynamics with both potentially positive and negative outcomes: *The promise of digital agility* on the one hand, and *the risk of digital fragility* on the other.

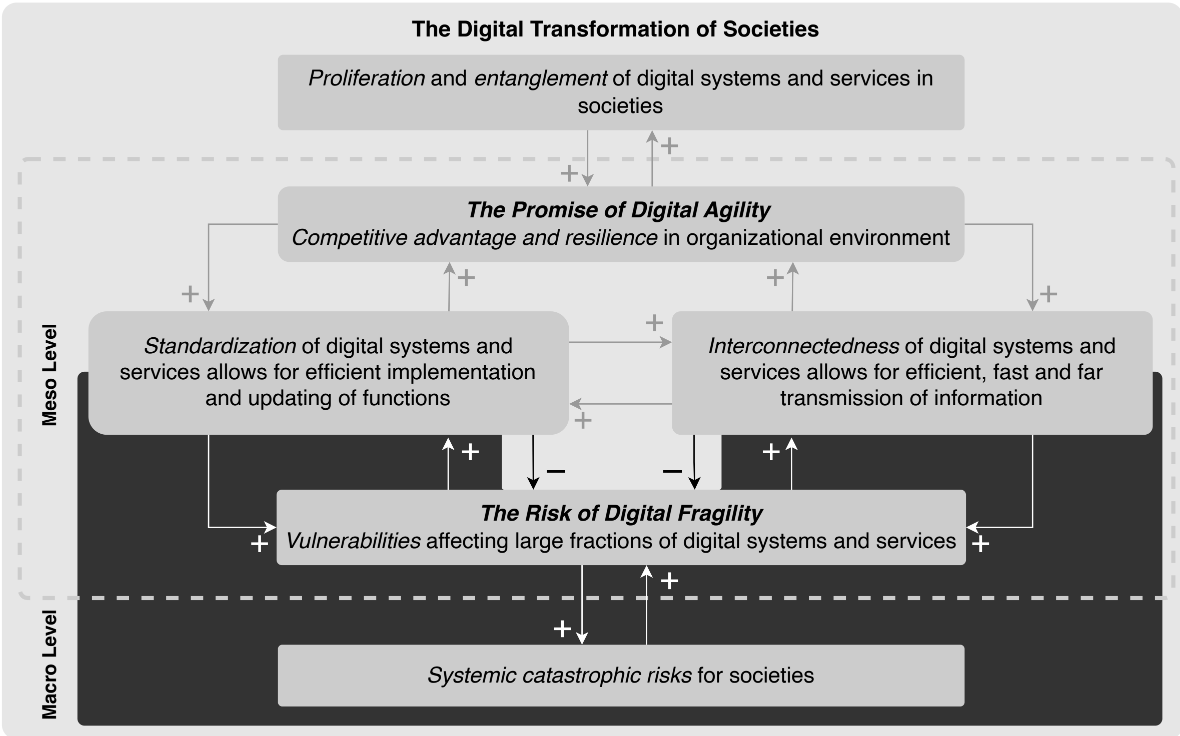

Figure 6: The Promise of Digital Agility and the Risk of Digital Fragility.

On the potentially bright side, the promise of digital agility refers to the meso-level dynamics where the characteristics of digital systems and services provide competitive advantage and resilience in organizational environments, creating reinforcing positive feedback loops that drive adoption. At the macro-level these dynamics become visible and are further reinforced by the proliferation and entanglement of digital systems and services in societies. The more digital systems and services there already are, the easier and more promising further digitalization becomes (cf. Metcalfe, 2013). Thus, we see the promise of digital agility as an underlying driver of the ongoing digital transformation of our societies.

On the dark side, we argue that the increasing digitalization of our societies comes at the cost of an increasing risk of digital fragility. By this we mean that the aforementioned characteristics also contribute to the creation of a meso-level dynamic, where vulnerabilities can affect large fractions of digital systems and services at the same time, which in turn contributes to macro-level systemic catastrophic risks. The larger the fraction of digital systems and services that can be affected simultaneously by a vulnerability, the greater the risk of catastrophic rather than manageable failure in situations when it counts (cf. Scheffer et al., 2012).

On the one hand, pressure towards standardization creates a more uniform landscape of technologies in the lower layers of the digital technology architecture (Yoo et al., 2010), making many digital systems susceptible to common vulnerabilities. For example, a bug in a low-level instruction set architecture can affect all systems using that architecture (Abu-Ghazaleh et al., 2019). Additionally, all digital systems are vulnerable to electricity loss. However, standardization can also help to close or even prevent vulnerabilities as it enables a faster distribution of fixes once vulnerabilities are discovered and can promote the use of best practices and the adoption of, ideally, provably safe systems. Nevertheless, this remediation process is not always perfect and can be hindered by situational constraints, as seen, for example, in the slow update cycles of many Android devices (Mahmoudi & Nadi, 2018).

On the other hand, increasingly interconnected digital systems and services can lead to the rapid spread of cyber contagions and disruptions. Vulnerabilities in large-scale digital platforms can have a global impact as demonstrated by outages at digital infrastructure providers affecting thousands if not millions of customers and their organizations around the globe (e.g., Satariano, 2021). In particular, due to the complex nature of how large networks are set up, even routine maintenance can potentially trigger cascading failures with critical consequences. For example, on October 4, 2021, routine maintenance at Meta caused a complete outage of its entire network for several hours, because its systems were interconnected in a way that allowed a maintenance operation to affect systems it should not have affected, while making recovery operations more difficult (Janardhan, 2021). On the flip side, interconnectedness also potentially allows for more rapid vulnerability fixing. For this to happen, however, situational and institutional constraints must allow this to become a priority (Dunn Cavelty, 2014).

This generalized conceptualization helped us to (1) better understand the dynamics we observed in our simulation experiments and also enabled us to (2) propose a tentative set of strategies for managing the risk of digital fragility that we identified.

First, our simulation experiments show how catastrophic failures in critical sectors of society can occur when a large fraction of digital systems and services are affected by a prolonged power outage. At some point, a tipping point is reached where the induced stress simply overwhelms the local adaptive capacity of the system (cf. Scheffer et al., 2012). These dynamics apply to various catastrophic scenarios, not just power outages. Loss of electricity is a significant vulnerability for digital and non-digital systems, but other potent vulnerabilities will emerge as our societies become more standardized and interconnected, making the risk of digital fragility a potentially extremely dangerous Achilles' heel for our societies (Graham, 2011; Manheim, 2020).

Second, drawing on the available literature on catastrophic risks and emergency preparedness (e.g., Stockton & Council, 2016, 2018), and the general notion of antifragility (Taleb, 2012), which describes generic strategies for overcoming fragility such as creating optionality and redundancy in the system (Derbyshire & Wright, 2014), we suggested a tentative set of seven promising strategies to manage the risk of digital fragility with the aim of identifying promising avenues for future research. An overview of the strategies is given in Table 3. A more detailed discussion is presented in Publication 3.

Overall, the key findings of the third inquiry illustrate how a responsible DSR discourse can be initiated by focusing on the development of a proof-of-problem for a societally relevant challenge of the digital age. The risk of digital fragility is a novel framing of a highly problematic, unintended consequence of the ongoing digital transformation of our societies that we are to ignore at our own peril. In this sense, this work should be seen as taking only the first steps on a long and winding road of responsible DSR discourse. Much more work is needed to better understand the risk of digital fragility and to find ways to manage it. The work we have done so far should provide fertile ground for such efforts. Given the strong interdependencies between this applied work and the meta-research on responsible DSR, this also represents a meaningful contribution to the development of the responsible DSR ecosystem in general. It is well understood that prototypical examples of good work are necessary for emerging paradigms to gain traction and recognition (Kuhn, 2012; Lee, 2000).

Table 3: Promising Strategies to Manage the Risk of Digital Fragility.

#	Description
1	*Modularity Through Protected Enclaves* Increasing network modularity reduces the impact of disruptions. Similarly, creating protected enclaves in critical infrastructure increases resilience. For example, loosely coupled microgrids reduce electricity grid interdependencies. Protected enclaves are critical for recovering from large-scale disruptions. IS researchers can assist in their design and development.
2	*Optionality Through Alternative Communication Systems and Power Sources* Creating optionality, such as alternative communication systems that can replace the Internet in an emergency, is essential to resilience. The development of diverse communication mechanisms such as wireless technology and mesh networks can help. In addition, in disaster scenarios, hybrid electric vehicles (HEVs) can serve as mobile power sources, coordinated by smartphone-based emergency systems, to enhance recovery efforts. Research to integrate HEVs and expand emergency power options is a promising avenue to contribute to disaster resilience.
3	*Redundancy Through Readily Available Replacement Parts* Redundancy and spare capacity are valuable in preparing for uncertainty. While they may reduce efficiency, they provide significant benefits during emergencies, as seen during the COVID-19 pandemic. In critical infrastructure, ensuring that SCADA systems can work with replaceable computer parts can improve redundancy. Research can explore innovative ways to secure spare parts for digital systems, drawing on the concept of the circular economy.
4	*Diversity Through Small-Scale Experimentation* Diversity in a system increases redundancy and resilience. Small-scale experimentation and open innovation initiatives can foster this diversity and mitigate the risks of digital fragility. IS research, building on existing lessons from open innovation, can be a leader in advancing novel strategies and approaches.
5	*Hormesis Through Bug Bounty Programs* Hormesis, in the context of planning for rare catastrophic events, involves seeking stressors to prepare organizations. Bug bounty programs (BBPs) are a cost-effective and secure way to do this. They use crowdsourcing and prizes to identify vulnerabilities in target systems, especially in critical sectors undergoing digital transformation. IS research can lead and scale such initiatives.

#	Description
6	*Better Preparedness Through Financial Incentives* Resilience bonds, which are similar to insurance but offer discounts on premiums when risk reduction projects are implemented according to probabilistic catastrophe models, make disaster preparedness more manageable for organizations. Mandating resilience bonds for critical infrastructure sectors can reduce costs and promote long-term resilience. Despite the potential, challenges remain in modeling critical infrastructure risks. IS research can help develop simulation models for practical implementation.
7	*Cooperative Response Through Legal Obligations* Legally binding agreements for critical infrastructure providers to work together during extreme disasters are essential. Mobile phone agreements for emergency calls could serve as a model. Extending this to ensure Internet access during disasters would support the dissemination of essential information. Studying coopetition strategies, such as those used during the COVID-19 pandemic, can help to establish effective cooperation mechanisms. Future IS research can contribute by designing and evaluating coordination methods and platforms for effective governance.

4.4 Synthesis

The main results presented in this dissertation highlight challenges and opportunities in the field of IS research, with a focus on ethics, responsible DSR, and the risk of digital fragility in the digital age. Collectively, they make significant progress toward the overarching goal of this dissertation: *To initiate and support the evolution of a responsible DSR ecosystem that can help implement and advance RI by addressing the grand challenges of the digital age.*

The first inquiry underscored the importance of integrating ethics into IS research. It identified two main challenges: the lack of skills and resources for ethical decision-making and value conflicts arising from diverse value systems. Opportunities identified include value reporting, problem prioritization to promote ethical awareness and improve research relevance, and the deliberate development of infrastructures to more systematically support ethical conduct and high impact research. Taken together, these findings highlight the need to systematically integrate ethical considerations into IS research to ensure responsible and relevant outcomes.

The second inquiry developed a framework for responsible DSR based on four interrelated paradigmatic challenges: ontological, epistemological, axiological, and methodological. Rather than providing specific answers on how to conduct responsible research, this framework focuses on clarifying the nature of its practical challenges. In doing so, it offers a flexible and integrative perspective on responsible DSR that can guide the IS research community in developing a research ecosystem that facilitates responsible and effective research practices and thus helps us reap the benefits, but avoid the potential downsides, of the ongoing digital transformation of our societies. To kick-start this process, initial recommendations for the implementation of such an ecosystem have also been derived and presented.

The third inquiry, building on the previous inquiries, initiated a prototypical responsible DSR discourse by exploring the risk of digital fragility in our increasingly digitalized societies. It revealed how the ongoing digital transformation of our societies, while promising digital agility, also produces vulnerabilities that contribute to systemic catastrophic risks.

These findings highlight the critical need to manage and mitigate the risk of digital fragility to ensure the resilience of our societies. They also illustrate the types of deep and relevant societal challenges that a responsible DSR ecosystem could help address.

In summary, these inquiries collectively contribute to a more systematic integration of ethics into IS research by outlining a vision for a responsible DSR ecosystem and taking first steps toward its implementation. However, more work is needed to make this vision a reality. This requires more and deliberate engagement with the paradigmatic challenges of responsible DSR. It also requires more work that directly addresses the societal issues we face in the digital age. Finally, it outlines how IS researchers can play a pivotal role in society if we choose to unite our efforts in a responsible DSR ecosystem. Such efforts will be critical to ensuring that digital technologies are a force for good and do not inadvertently create serious risks. If the long-term well-being and success of society are our research questions, responsible DSR is the inquiry to find the answers.

Discussion

5.1 Limitations

I want to acknowledge and discuss three key limitations of this work: (1) the inherent selectivity of its focus and engagement, (2) the limited degree of community participation in the design and execution of the inquiries, and (3) the relatively nascent state of the findings.

First, in acknowledging a critical pragmatist stance, I want to be frank and emphasize the inherent selectivity of this work. All three of the inquiries that make up this dissertation are necessarily highly selective in focusing on certain aspects and deemphasizing others, which raises the question of how well reasoned and transparently documented these choices are. For example, looking at Inquiry 1, one might ask whether the perspectives we have discussed provide a comprehensive view of the diverse range of potentially relevant perspectives. Have we broadened the focus of our inquiry enough? As discussed when I introduced critical pragmatism in the theoretical foundations section, there is no way for us to fully justify our choices here. The best we could do was to document our choices transparently in order to allow for and encourage criticism. The same can be said for the other two inquiries. I must acknowledge that none of the choices regarding their focus and scope can be fully justified, but at least I have tried to document them transparently to allow for critical discourse. In this sense, all of the findings should be seen as the endpoints of serious and rigorous investigations, but at the same time only as starting points for larger conversations.

Second, building on the aforementioned inherent selectivity of this work, I must mention as a potential limitation the limited degree of community participation in the design and execution of the inquiries. The goal of this dissertation is to initiate and support the development of a responsible DSR ecosystem, which necessarily implies the need for community participation. However, while I have considered and included many voices and perspectives from the IS research community in this work, their actual involvement in the design and execution of this work has been limited. Rather than engaging in participatory co-creation with community members, it could be argued that I ended up treating them more or less as interesting research subjects to be studied and theorized about. This means that the findings of this dissertation cannot be taken as representative of the perspectives of the IS research community at large, but should rather be seen as the personal perspectives and visions of the author of this dissertation and his respective co-authors. Thus, more critical discourse and participatory engagement will be needed to identify possible differences in understanding between relevant parties and to make progress towards a vibrant and aligned responsible DSR ecosystem. Ultimately, this dissertation could only take the very first steps on what will ultimately be a long and winding road that will require further commitment and hard work to travel.

Third, the relatively nascent state of the findings means that it is difficult, if not impossible, to objectively assess the quality and contribution of this dissertation as a whole. From a pragmatist perspective, what is needed to reach a final judgment is a comprehensive transformation of the problematic situation into one that is more or less resolved (Dewey, 1938). Until this point of resolution is reached, all findings must be regarded as highly tentative (Dewey, 1941). However, given the ambition of this dissertation, I cannot in any way claim to have come close to a state of resolution. Quite the opposite, in fact. If this dissertation is well received and taken seriously, it is likely to generate a good deal of controversy and critical discussion about the ends of inquiry and what this should mean for the identity and direction of the IS research community.

This situation leaves us with the question of how this dissertation should be judged, if not by its demonstrated ability to resolve a problematic situation? In the spirit of critical pragmatism, I suggest that the answer lies in assessing its critical potential. As our world becomes more complex in the digital age, comprehensive resolutions to problematic situations become more elusive. As Ulrich (1994) has pointed out, in such situations, contributions that help us see problematic situations from a new perspective and in a different light often turn out to be the most relevant ones in retrospect. In this sense, meaningful and reasonable critique can be seen as the engine that drives progress toward eventual resolution and a useful ideal against which to measure this dissertation. So the question becomes: Does this dissertation make a reasonable case for a new and meaningful way of looking at responsible DSR that can expand and inform our thinking? I certainly hope so, and invite the reader to make up his or her own mind.

5.2 Contributions

5.2.1 Contributions to Theory

I highlight four main theoretical contributions of this dissertation as a whole: (1) the development of a critical pragmatist perspective on responsible DSR that integrates elements of Dewey's (1938) theory of inquiry with Ulrich's (2006, 2016) proposal for a critical turn in pragmatism in a novel and coherent way, (2) presents a comprehensive, integrative framework for understanding the key practical challenges for RI in the digital age, and (3) makes progress toward resolving the paradigmatic divide in IS research by offering an integrative "super-methodology" that can facilitate the integration of the broad intellectual diversity of the IS research community in pursuit of common interests; (4) the formulation of a novel proof-of-problem for IS research that expands our understanding of the potential dark side of digital technologies and holds great promise for promoting more societally relevant IS research.

First, the paradigmatic framework for responsible DSR (see Publication 2) builds on and integrates the work of Dewey (1938) and Ulrich (2006, 2016) in a novel and coherent way. Because Ulrich's (2006, 2016) focus was primarily on making the case for a critical turn in pragmatism and elaborating on how his own work on CSH (Ulrich & Reynolds, 2010) can inform and support such an effort, there was still a gap in terms of integrating his proposal into a comprehensive, coherent, and self-sufficient critical pragmatist perspective. We argue that the paradigmatic framework for responsible DSR helps to fill this gap. Building not only on Ulrich's work, but also on that of Dewey and many other scholars from a variety of fields, not least the practical insights and experiences of IS researchers, I have developed a comprehensive account of critical pragmatism as a theory of inquiry that conceptualizes and distills its ontological, epistemological, axiological, and methodological challenges. In doing so, I have extended the perspective offered by Ulrich (2006), who focused on critical pragmatism, not as a general theory of inquiry, but as a new approach to professional and business ethics. While assessing the merits of this reframing of the critical pragmatist perspective certainly requires further debate and critical discourse, I suggest that for this very reason this work has considerable critical potential and thus represents a significant contribution to the academic discourse.

Second, building on the previous point, I argue that much of the critical potential of the framework for responsible DSR lies in the fact that it provides a novel, comprehensive, and actionable account of key practical challenges for RI in the digital age that can serve as a reference point for comparing and integrating competing visions for RI (cf. Timmermans & Blok, 2021). The focus on the ontological, epistemological, axiological, and methodological dimensions of responsible DSR paints a comprehensive but not overwhelming picture of four key practical challenges that need to be addressed in RI practice. This work thus responds to calls for greater conceptual clarity and less naive perspectives on RI that are better able to address the practice dimension of RI (e.g., Blok, 2023; Blok & Lemmens, 2015; Novitzky et al., 2020; Timmermans & Blok, 2021). In particular, the framework has already been shown to lend itself to the systematic derivation of practical recommendations that can inform the implementation of RI. For example, this work identified value conflicts as a potential barrier to the successful implementation of RI, and responded with specific recommendations for establishing an open, responsible DSR ecosystem that facilitates polycentric, multi-level cooperation and value co-creation, and allows for the negotiation of value conflicts as they arise. In doing so, this work provides additional conceptual underpinnings for recent RI research efforts that have begun to embrace similar visions, for example, in the form of social labs that study and promote RI practices through a hermeneutic action research approach (Blok, 2023; Timmermans et al., 2020).

Third, considering the theoretical implications for IS research, the critical potential of the framework for responsible DSR lies in its ability to act as a "super-methodology" that can facilitate the integration of the broad intellectual diversity of the IS research community (e.g., Rai, 2017a; Tarafdar et al., 2022) in the pursuit of common interests, such as strategic and systematic programs of high-impact research (e.g., Gable, 2020; Nunamaker et al., 2017; Ram & Goes, 2021; Winter & Butler, 2011). This argument parallels that of Morgan (2007), who described how the so-called "metaphysical paradigm," the hitherto hegemonic idea that research must be located within a set of competing and incommensurable research paradigms defined by their metaphysical assumptions (e.g., Post-Positivism vs. Constructivism vs. Critical Theory), could be replaced by a pragmatic approach as an "*integrated methodology* for the social sciences" (Morgan, 2007, p. 73). In essence, as a contextualist philosophy, pragmatism does not assume incommensurable differences in the world, but rather advocates for open-minded and open-ended inquiry (Dewey, 1938), and thus can be used with all types of research methods, be they qualitative, quantitative, design, or mixed methods (Morgan, 2014). I suggest, therefore, that the pragmatic framework for responsible DSR can serve as a grounding lightning rod for the diverse intellectual capacities and interests of the IS research community and beyond. As such, this work not only helps to bridge the longstanding and increasingly untenable divide between DSR and BSR (e.g., Alter, 2012; Herwix & Rosenkranz, 2018; McKay et al., 2012), but also responds to recent calls for greater support for strategic and systematic programs of high-impact research (e.g., Desouza & Dawson, 2023; Gable, 2020; Nunamaker et al., 2017; Ram & Goes, 2021; Winter & Butler, 2011).

Fourth, the third inquiry proposed a novel proof-of-problem by presenting the risk of digital fragility. This problem represents a great opportunity for IS research to make societally relevant contributions. Given the potentially catastrophic consequences of the risk of digital fragility on a societal scale, it can be argued that significant amounts of resources deserve to be allocated to work on this problem (Herwix & Haj-Bolouri, 2021). In short, the larger the scale of the problem, the more tractable a resolution, and the more neglected the current work on it, the greater the expected value of working on it. Thus, given the current neglect of the topic and the relevant expertise of the IS research community to work on it, directing further resources to work on this problem seems to be a great opportunity to live up to our ideals and create more societal value with our work (Desouza & Dawson, 2023; Herwix & Haj-Bolouri, 2020).

5.2.2 Contributions to Practice

I highlight two major contributions of this dissertation to practice: (1) all of the inquiries, individually and collectively, have helped to advance our understanding of RI practice with direct practical implications, and (2) the MLDM in particular provides helpful guidance not only for individual researchers and the IS research community, but also for all practitioners seeking to implement RI.

First, I summarize how the findings and frameworks developed in each inquiry provide practical insights and guidance for RI practitioners:

1. The first inquiry identified key challenges and opportunities for ethics in IS research and practice, providing actionable insights for research practitioners. Practical implications highlighted include the need for research organizations to allocate resources and support for ethical decision-making, the benefits of promoting value reporting to increase transparency, and the value of prioritizing research problems with ethical considerations in mind.

2. The second inquiry presented a paradigmatic framework for responsible DSR that can be used by researchers, as well as other RI practitioners, to reflect on and improve their work practices. In addition to these implications at the individual level, the systemic implications of the framework for the development of a responsible DSR ecosystem were also discussed in the form of four specific recommendations.

3. The third inquiry explored the risk of digital fragility in the digital age and provides strategies for managing this risk, offering practical guidance for mitigating the systemic impact of vulnerabilities in digital systems. These strategies, such as modularity, optionality, redundancy, and diversity, can be studied by IS researchers, implemented by critical infrastructure practitioners, and supported by society at large to enhance resilience and reduce the risk of digital fragility.

Taken together, these inquiries develop a vision of a responsible DSR ecosystem that seeks to inspire and guide RI practitioners. Even if only the first steps have been taken in this direction, the realization that it may indeed be a direction worth taking can sometimes be the hardest to come by. In this sense, a collective, practical contribution of this dissertation lies in its visionary character, which can open new horizons and stimulate critical reflection for IS researchers and RI practitioners alike.

Second, I highlight the Multi-Layered Development Model for responsible DSR (MLDM), which is presented in detail in the third chapter of this dissertation, as a particularly promising contribution to RI practice. In essence, the MLDM represents a departure from traditional DSR research processes, emphasizing collaboration and peer review across the responsible DSR ecosystem as part of the research process in response to the rapidly evolving technological landscape and its associated risks. As such, it recognizes that responsible DSR is a discursive practice that is not limited to individual research projects, but extends to an entire research ecosystem. It introduces a stage-gate-based development model with four layers[1] and four stages[2] to ensure an iterative progression of responsible DSR. The MLDM acknowledges that implementing a stage-gate-based model would place high demands on responsible DSR practitioners. However, it argues that the potential downsides of unrestrained technology development and the benefits in terms of better problem prioritization and increased collective productivity justify the investment. It thus makes a significant practical contribution by proposing a novel model for organizing responsible DSR that can inform not only the envisioned responsible DSR ecosystem, but also RI practitioners and organizations more broadly.

5.2.3 Contributions to Policy

This dissertation makes two main contributions to policymaking that are noteworthy: (1) four strategic recommendations to guide policymaking in establishing responsible DSR ecosystems, and (2) seven strategies for mitigating the risk of digital fragility to inform critical infrastructure policymaking.

First, the second inquiry developed four strategic recommendations for developing a responsible DSR ecosystem that can help to rethink established research and innovation policies in light of the growing demand for RI in the digital age. Some illustrative examples include, but are not limited to:

- *Establish Systematic Boundary Discourse:* Incentives and funding mechanisms can be structured to reward projects that demonstrate transparency in boundary judgments and to encourage the development of tools and methodologies for systematic boundary critique.

- *Support Responsible Multi-Layered Development*: Policymakers can promote collective responsibility in the responsible DSR ecosystem by enabling collaboration across projects. This can be achieved through grant programs, RI networks, and RI platforms that encourage inter-, multi-, and transdisciplinary coordination, for example, by evaluating projects not only on their individual merits but also on their contribution to the overall responsible DSR ecosystem.

- *Facilitate Polycentric, Multi-level Cooperation and Value Co-creation:* Policymakers can create governance frameworks that facilitate and incentivize polycentric, multi-level cooperation. They can create forums for relevant parties from different levels and sectors (e.g., academia, industry, civil society, government) to come together and share insights and challenges, leading to the initiation of collaborative projects that promote value co-creation.

[1]The four layers represent an expanding focus as problematic situations move closer to resolution: (1) framing the problematic situation establishes a *proof-of-problem*; (2) prototyping potential solutions produces a *proof-of-concept*; (3) evaluating value propositions justifies a *proof-of-value*; and (4) responsibly deploying effective solutions sustains a *proof-of-use*. These layers are additive, with insights from lower layers informing and influencing higher layers. Ongoing open discourse at each layer is essential to facilitate effective and responsible coordination within the responsible DSR ecosystem.

[2]The four stages act as checks and balances to ensure responsible progress toward resolution: (1) the *Problem Stage* can be passed with a justified proof-of-problem; (2) the *Concept Stage* can be passed with a proof-of-problem and a working proof-of-concept; (3) the *Value Stage* can be passed with a proof-of-problem, proof-of-concept, and verified proof-of-value; and (4) the *Use Stage* is an open-ended stage focused on maintaining proofs at all levels while ensuring the responsible deployment of a solution. These stages are designed to ensure iterative critical reflection and peer review throughout responsible DSR processes, which can enhance responsibility and address challenges related to the societal impacts of digital systems and technologies.

- *Develop Pragmatism, Reflexivity, Sensitivity, and Anticipation:* Policymakers can support the development of the necessary skills for responsible DSR. This can be done by providing training programs, educational incentives, and resources for researchers and practitioners to engage in lifelong learning. Policies can also promote reflexivity, sensitivity, and anticipation by encouraging self-assessment, continuous improvement, and adaptability in responsible DSR processes.

In summary, this work calls on policymakers to play a crucial role in establishing a responsible DSR ecosystem by incorporating its findings into governance frameworks. This includes providing funding, creating incentives, and fostering a collaborative culture consistent with the principles of responsible DSR as outlined in the recommendations.

Second, the third inquiry developed seven strategies for mitigating the risk of digital fragility that can directly inform critical infrastructure policymaking. In particular, I argue that this work provides appropriate starting points for more comprehensive policymaking processes that would further explore the implications of each proposal in detail. Recognizing this limitation, I suggest the following key takeaways as possible starting points for policymakers seeking to mitigate the risk of digital fragility:

- *Promote Modularity and Protected Enclaves:* Encourage the development of protected enclaves within critical infrastructure to increase network modularity and reduce interdependencies.

- *Foster Optionality:* Incentivize the creation of alternative communication systems and power sources to ensure resilience, especially during emergencies.

- *Ensure Availability of Spare Parts*: Ensure the ready availability of spare parts for critical infrastructure to improve redundancy, even if it reduces efficiency.

- *Encourage Diversity Through Experimentation*: Promote diversity in critical infrastructure by supporting open innovation and small-scale experimentation to increase redundancy and resilience.

- *Scale Bug Bounty Programs*: Endorse and scale bug bounty programs to identify vulnerabilities in digital systems in critical sectors.

- *Mandate Resilience Bonds*: Establish mandatory resilience bonds for critical infrastructure sectors that incentivize risk reduction and promote long-term resilience.

- *Establish Legal Obligations for Cooperation*: Create legally binding agreements for critical infrastructure providers to cooperate during extreme disasters.

In conclusion, these strategies provide concrete and actionable starting points for critical infrastructure policymaking to reduce the risk of digital fragility. By considering these strategies and creating the necessary policy frameworks, governments and organizations can improve the resilience of their critical infrastructure in the face of severe disruptions and unforeseen events.

5.3 Future Research

The key findings of this dissertation open up three major directions for future work: (1) the work on the risk of digital fragility has initiated a responsible DSR discourse that calls for further engagement; (2) the work on the paradigmatic foundations of responsible

DSR calls for further reflection and refinement in light of its ongoing application; (3) this dissertation as a whole calls for the development of infrastructure and institutions that better support and facilitate responsible DSR.

First, the work on the risk of digital fragility shows how the digital transformation of our societies can be associated with serious systemic catastrophic risks. However, it only scratches the surface of this phenomenon by demonstrating its general plausibility. Further research is needed to better understand the mechanisms at play. For example, while we have explored the impact of digital transformation on the resilience of the U.S. food system in two CEL scenarios, there are certainly other scenarios and sectors where the risk of digital fragility could play an important role. In particular, the catastrophic failure of Meta's entire network infrastructure on October 4, 2021 (see Janardhan, 2021) could be an interesting case to explore how the standardization and interconnectedness of digital infrastructure components may have created an inherently fragile system that was bound to experience a catastrophic failure at some point. This could be done through additional simulation models that would help us better understand the dynamics of digital fragility at different scales and in different scenarios. In addition, more research is needed to further investigate strategies for mitigating the risk of digital fragility. While we have already proposed seven promising strategies, we have yet to assess their expected value in terms of cost-effectiveness and to explore potential unintended consequences in sufficient detail. This opens up promising opportunities for future work to help identify and prioritize well-considered, integrated policy mixes that can mitigate the risk of digital fragility at a reasonable cost.

Second, the paradigmatic framework for responsible DSR provides a promising foundation for future research on responsible DSR. On the one hand, it can inform the design and implementation of applied responsible DSR programs and projects. On the other hand, future research can reflect on and refine the framework itself in light of the experience gained through its application. Ideally, these two strands of future research will be productively intertwined in a virtuous cycle of research, so that experiences from the applied research strand will feed directly into the reflective research strand, which in turn will provide useful guidance to the applied research strand. With this dissertation, I have attempted to start such a virtuous cycle by initiating an applied responsible DSR discourse that draws on and instantiates the framework for responsible DSR to address a real-world problem. Future research is called upon to weave the lessons to be learned from this endeavor back into the ongoing development of the responsible DSR framework, and to continue the virtuous research cycle through further applications. The goal is to iteratively develop a vibrant and thriving responsible DSR ecosystem over time. Notably, this vision is somewhat similar to Timmermans et al.'s (2020) proposal to use a hermeneutic action research approach as the basis for creating social labs as real-world incubators for studying and promoting RI practices in society. Specifically, their proposal also calls for a circular approach to studying and applying practices that support RI. Given this similarity in interest and methodological orientation, future research is encouraged to further explore the potential synergies between these two approaches in more detail.

Third, this dissertation as a whole argues that more research should focus on developing infrastructures and institutions that are better suited to the demands of responsible DSR. Of course, since this work has focused mainly on clarifying the fundamental challenges that such infrastructures and institutions must help to address, it can only take us the very first steps on a long and winding road towards more responsible practice. I must therefore call on future research to follow in my footsteps and to seriously consider how each of us, and all of us together, can contribute to a more responsible DSR ecosystem. In doing so, I stand in a long tradition of scholars, such as Ulrich (1994), Winter and Butler

(2011), Nunamaker et al. (2017), Morana et al. (2018), Gable (2020), and Desouza and Dawson (2023), all of whom, in their own ways, have urged IS researchers to take greater responsibility and work together decisively for the benefit of our collective future. Beyond these previous efforts, however, this work provides a novel, multi-grounded theoretical foundation that can rigorously support and guide such future development efforts. As such, future research should now be better positioned to actually respond to the resounding calls for greater responsibility.

Conclusion

In this dissertation, I have embarked on a journey to initiate and support the development of a responsible DSR ecosystem that can help implement and advance RI by addressing the grand challenges of the digital age in a responsible and productive manner. This journey began with the recognition of the transformative impact of digital technologies and the dual nature of their potential—to bring unprecedented value to our societies, but also to pose significant threats if not managed responsibly. We now live in a world where advanced AI technologies, including the potential for AGI, raise existential questions. The need for RI has never been more apparent.

In response, I have explored the complexities of implementing and advancing RI, only to discover that our understanding of how to translate RI into practical action is still in its infancy. Given the rapid pace of technological progress and its profound impact, implementing RI in the digital age is a daunting challenge. At its core, RI requires us to consider what kind of future we want to create, and then to question our motives and values in relation to those of others in order to find responsible ways to ensure that the future we arrive at remains desirable for all concerned, while recognizing the inherent selectivity and uncertainty involved in this task.

Looking at the IS research community, I realized that a major paradigm shift is needed to unleash its full potential in this regard. We need to transcend the conventional boundaries between BSR and DSR that have divided the IS research community for the past several decades if we are to implement and advance RI by addressing the grand challenges of the digital age in a responsible and productive manner. A responsible DSR ecosystem, based on an open-minded and open-ended pragmatist understanding of research, where inquiry is guided by contextual needs rather than preconceived notions of methods and outcomes, is needed to bridge the divides and unite IS research in a responsible DSR ecosystem.

This dissertation has presented an argument and framework for such a responsible DSR ecosystem, has characterized it in terms of ontological, epistemological, axiological, and methodological challenges, and demonstrated its practical applicability in the development of a novel research problem with potentially great societal impact. In concluding this dissertation, however, I must acknowledge that this work only scratches the surface of the complex landscape of implementing and advancing RI in the digital age. It serves as a starting point for an ongoing, interactive, and dialogical journey. RI calls for a transparent and participatory process in which societal actors and innovators respond to one another to ensure the ethical acceptability, sustainability, and societal desirability of the innovation process. The same requirements apply to this work.

In the digital age, the road ahead is filled with challenges and opportunities. This dissertation invites others to join me on the journey toward a responsible DSR ecosystem that can help implement and advance RI by addressing the grand challenges of the digital age in a responsible and productive manner. Only through such deliberate collective efforts can we help shape the future in a way that is consistent with our values, ethics, and societal well-being. This is our responsibility, and this dissertation is but a stepping stone.

References

Abu-Ghazaleh, N., Ponomarev, D., & Evtyushkin, D. (2019). How the spectre and meltdown hacks really worked. *IEEE Spectrum*, *56*(3), 42–49. https://doi.org/10.1109/MSPEC.2019.8651934

Ackoff, R. L. (1974). *Redesigning the future: a systems approach to societal problems*. New York, Wiley. Retrieved July 11, 2021, from http://archive.org/details/redesigningfutur00russ

ACM Code 2018 Task Force. (2018). Acm Code of Ethics and Professional Conduct. Retrieved September 22, 2021, from https://dora.dmu.ac.uk/bitstream/handle/2086/16422/acm-code-of-ethics-and-professional-conduct.pdf

Allen, L., O'Connell, A., & Kiermer, V. (2019). How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learned Publishing*, *32*(1), 71–74. https://doi.org/10.1002/leap.1210

Alter, S. (2012). Long Live Design Science Research! .... And Remind Me again about Whether It Is a New Research Paradigm or a Rationale of Last Resort for Worthwhile Research that Doesn't Fit under Any Other Umbrella. *ICIS 2012 Proceedings*, 6.

Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures*, *46*, 23–40. https://doi.org/10.1016/j.futures.2012.10.003

Andersen, B. P., Miller, M., & Vervaeke, J. (2022). Predictive processing and relevance realization: Exploring convergent solutions to the frame problem. *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-022-09850-6

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Baham, C., & Hirschheim, R. (2022). Issues, challenges, and a proposed theoretical core of agile software development research. *Information Systems Journal*, *32*(1), 103–129. https://doi.org/10.1111/isj.12336

Barrett, W., & Aiken, H. D. (Eds.). (1962). *Philosophy in the twentieth century : an anthology* (Vol. 1). Random House. Retrieved September 20, 2023, from http://archive.org/details/philosophyintwen0000unse

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design Science Research Contributions: Finding a Balance between Artifact and Theory. *Journal of the Association for Information Systems*, *19*(5), 358–376. https://doi.org/10.17705/1jais.00495

Baskerville, R., Kaul, M., & Storey, V. C. (2015). Genres of Inquiry in Design-Science Research: Justification and Evaluation of Knowledge Production. *MIS Quarterly*, *39*(3), 541–564.

Baskerville, R., & Wood-Harper, A. (1998). Diversity in information systems action research methods. *European Journal of Information Systems*, 18.

Baskerville, R., Lyytinen, K., Sambamurthy, V., & Straub, D. (2011). A response to the design-oriented information systems research memorandum. *European Journal of Information Systems*, *20*. https://doi.org/10.1057/ejis.2010.56

Baskerville, R., & Pries-Heje, J. (2010). Explanatory design theory. *Business & Information Systems Engineering*, *2*(5), 271–282.

Bengio, Y. (2023, May 23). How Rogue AIs may Arise. Retrieved May 24, 2023, from https://yoshuabengio.org/2023/05/22/how-rogue-ais-may-arise/

Berti, M., Jarvis, W., Nikolova, N., & Pitsis, A. (2021). Embodied Phronetic Pedagogy: Cultivating Ethical and Moral Capabilities in Postgraduate Business Students. *Academy of Management Learning & Education*, *20*(1).

Bessant, J. (2013). Innovation in the Twenty-First Century. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible innovation: Managing the responsible emergence of science and innovation in society* (1st ed.). Wiley.

Blok, V. (Ed.). (2023). *Putting Responsible Research and Innovation into Practice: A Multi-Stakeholder Approach* (Vol. 40). Springer International Publishing. https://doi.org/10.1007/978-3-031-14710-4

Blok, V., & Lemmens, P. (2015). The Emerging Concept of Responsible Innovation. Three Reasons Why It Is Questionable and Calls for a Radical Transformation of the Concept of Innovation. In B.-J. Koops, I. Oosterlaken, H. Romijn, T. Swierstra, & J. van den Hoven (Eds.), *Responsible innovation 2: Concepts, approaches, and applications* (pp. 19–35). Springer International Publishing. https://doi.org/10.1007/978-3-319-17308-5_2

Boell, S. K., & Cecez-Kecmanovic, D. (2014a). A hermeneutic approach for conducting literature reviews and literature searches. *Communications of the Association for Information Systems*, *34*(1), 257–286.

Boell, S. K., & Cecez-Kecmanovic, D. (2014b). On being 'systematic' in literature reviews in IS. *Journal of Information Technology*, *30*(2), 161–173. https://doi.org/10.1057/jit.2014.26

Boh, W., Constantinides, P., Padmanabhan, B., & Viswanathan, S. (2023). Building Digital Resilience against Major Shocks. *Management Information Systems Quarterly*, *47*(1), 343–360.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Briggs, R. O., & Schwabe, G. (2011). On Expanding the Scope of Design Science in IS Research. In H. Jain, A. P. Sinha, & P. Vitharana (Eds.), *Service-oriented perspectives in design science research* (pp. 92–106). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20633-7_7

Brown, M. J. (2012). John Dewey's Logic of Science. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, *2*(2), 258–306. https://doi.org/10.1086/666843

Brundage, M., & Guston, D. H. (2019). Understanding the movement (s) for responsible innovation. In R. von Schomberg & J. Hankins (Eds.), *International handbook on responsible innovation. a global resource* (pp. 102–121). Edward Elgar Publishing Limited.

Bryant, A. (2017). *Grounded Theory and Grounded Theorizing: Pragmatism in Research Practice*. Oxford University Press.

Brynjolfsson, E., & McAfee, A. (2012, January 23). *Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*. Digital Frontier Press.

Buhmann, A., & Fieseler, C. (2021). Towards a deliberative framework for responsible innovation in artificial intelligence. *Technology in Society*, *64*, 101475. https://doi.org/10.1016/j.techsoc.2020.101475

Castells, M. (2011). *The Rise of the Network Society* (2. ed. with a new preface, [reprint]). Wiley-Blackwell.

Center for AI Safety. (2023). *Statement on AI Risk*. Retrieved May 30, 2023, from https://www.safe.ai/statement-on-ai-risk

Chen, W., & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information Systems Journal*, *14*(3), 197–235. https://doi.org/10.1111/j.1365-2575.2004.00173.x

Christian, B. (2020, October 6). *The Alignment Problem: Machine Learning and Human Values* (1st edition). W. W. Norton & Company.

Churchman, C. W. (1967). Guest Editorial: Wicked Problems. *Management Science*, *14*(4), B-141-B–146. https://doi.org/10.1287/mnsc.14.4.B141

Churchman, C. W. (1995). Ethics and science. *Systems Research*, *12*(4), 267–271. https://doi.org/10.1002/sres.3850120406

Churchman, C. W. (1984). *The systems approach* (Rev. and updated, 2. Laurel print). Laurel Book.

Collingridge, D. (1980). *The social control of technology*. Pinter.

Consilience Papers. (2022, June 26). Technology is Not Values Neutral: Ending the Reign of Nihilistic Design. Retrieved January 30, 2023, from https://consilienceproject.org/technology-is-not-values-neutral/

Cotra, A. (2022). Without specific countermeasures, the easiest path to transformative AI likely leads to AI takeover. Retrieved August 25, 2023, from https://www.lesswrong.com/posts/pRkFkzwKZ2zfa3R6H/without-specific-countermeasures-the-easiest-path-to

Cross, N. (1993). Science and design methodology: A review. *Research in engineering design*, *5*(2), 63–69.

Dalsgaard, P. (2014). Pragmatism and Design Thinking. *International Journal of Design*, *8*(1), 143–155.

Davison, R. M., Majchrzak, A., Hardin, A., & Ravishankar, M.-N. (2023). Special issue on responsible IS research for a better world. *Information Systems Journal*, *33*(1), 1–7. https://doi.org/10.1111/isj.12405

Deng, Q., & Ji, S. (2018). A Review of Design Science Research in Information Systems: Concept, Process, Outcome, and Evaluation. *Pacific Asia Journal of the AIS*, *10*(1), 1–36. https://doi.org/10.17705/1PAIS.10101

Derbyshire, J., & Wright, G. (2014). Preparing for the future: Development of an 'antifragile' methodology that complements scenario planning by omitting causation. *Technological Forecasting and Social Change*, *82*, 215–225. https://doi.org/10.1016/j.techfore.2013.07.001

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The Digital Platform: A Research Agenda. *Journal of Information Technology*, *33*(2), 124–135. https://doi.org/10.1057/s41265-016-0033-3

Desouza, K. C., & Dawson, G. S. (2023). Doing strategic information systems research for public value. *The Journal of Strategic Information Systems*, *32*(4), 101805. https://doi.org/10.1016/j.jsis.2023.101805

Dewey, J. (1938). *Logic: The Theory of Inquiry*. Henry Holt; Company, INC.

Dewey, J. (1941). Propositions, Warranted Assertibility, and Truth. *The Journal of Philosophy*, *38*(7), 169–186. https://doi.org/10.2307/2017978

Dickson, G. W., Benbasat, I., & King, W. R. (1982). The MIS area: Problems, challenges, and opportunities. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *14*(1), 7–12. https://doi.org/10.1145/1017702.1017705

Doherty, N. F., Coombs, C. R., & Loan-Clarke, J. (2006). A re-conceptualization of the interpretive flexibility of information technologies: Redressing the balance between the social and the technical. *European Journal of Information Systems*, *15*(6), 569–582. https://doi.org/10.1057/palgrave.ejis.3000653

Dong, J. Q. (2022). Using Simulation in Information Systems Research. *Journal of the Association for Information Systems*, *23*(2), 408–417. https://doi.org/10.17705/1jais.00743

Doyle, C., & Luczak-Roesch, M. (2019). We need the open artefact: Design Science as a pathway to Open Science in Information Systems research. *Proceedings of the 14th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2019*.

Drechsler, A., & Hevner, A. (2022). Knowledge Paths in Design Science Research. *Foundations and Trends® in Information Systems*, *6*(3), 171–243. https://doi.org/10.1561/2900000028

Dunn Cavelty, M. (2014). Breaking the Cyber-Security Dilemma: Aligning Security Needs and Removing Vulnerabilities. *Science and Engineering Ethics*, *20*(3), 701–715. https://doi.org/10.1007/s11948-014-9551-y

Eating your own dog food. (2023, October 5). In *Wikipedia*. Retrieved October 24, 2023, from https://en.wikipedia.org/w/index.php?title=Eating_your_own_dog_food&oldid=1178791117

Fang, Y., Lim, K., Qian, Y., & Feng, B. (2018). System Dynamics Modeling for Information Systems Research: Theory of Development and Practical Application. *Management Information Systems Quarterly*, *42*(4), 1303–1329.

Fu, D. Y., & Hughey, J. J. (2019). Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *eLife*, *8*, e52646. https://doi.org/10.7554/eLife.52646

Fuller, R. B. (1957). A comprehensive anticipatory design science. *Royal Architectural Institute of Canada*, *34*(8), 357–361.

Fuller, R. B. (1969). *Operating Manual For Spaceship Earth*. Estate of R. Buckminster Fuller.

Gable, G. G. (2020). Viewpoint: Information systems research strategy. *The Journal of Strategic Information Systems*, *29*(2), 101620. https://doi.org/10.1016/j.jsis.2020.101620

Galletta, D. F., Bjørn-Andersen, N., Leidner, D. E., Markus, M. L., McLean, E. R., Straub, D., & Wetherbe, J. (2019). If Practice Makes Perfect, Where do we Stand? *Communications of the Association for Information Systems*, 39–64. https://doi.org/10.17705/1CAIS.04503

Galliers, R. (2003). Change as Crisis or Growth? Toward a Trans-disciplinary View of Information Systems as a Field of Study: A Response to Benbasat and Zmud's Call for Returning to the IT Artifact. *Journal of the Association for Information Systems*, *4*(1), 337–352. https://doi.org/10.17705/1jais.00040

Genus, A., & Stirling, A. (2018). Collingridge and the dilemma of control: Towards responsible and accountable innovation. *Research Policy*, *47*(1), 61–69. https://doi.org/10.1016/j.respol.2017.09.012

Giampietro, M., & Mayumi, K. (2018). Unraveling the Complexity of the Jevons Paradox: The Link Between Innovation, Efficiency, and Sustainability. *Frontiers in Energy*

Research, *6*. Retrieved September 13, 2023, from https://www.frontiersin.org/articles/10.3389/fenrg.2018.00026

Ginige, T., De Silva, L., Walisadeera, A., & Ginige, A. (2018). Extending DSR with Sub Cycles to Develop a Digital Knowledge Ecosystem for Coordinating Agriculture Domain in Developing Countries. In S. Chatterjee, K. Dutta, & R. P. Sundarraj (Eds.), *Designing for a digital and globalized world* (pp. 268–282). Springer International Publishing.

Godfrey-Smith, P. (2002). Dewey on Naturalism, Realism and Science. *Philosophy of Science*, *69*(S3), S25–S35. https://doi.org/10.1086/341765

Goldkuhl, G. (2011). Design research in search for a paradigm: Pragmatism is the answer.

Goldkuhl, G., & Cronholm, S. (2010). Adding Theoretical Grounding to Grounded Theory: Toward Multi-Grounded Theory. *International Journal of Qualitative Methods*, *9*(2), 187–205. https://doi.org/10.1177/160940691000900205

Goldkuhl, G., & Cronholm, S. (2018). Reflection/commentary on a Past Article: "Adding Theoretical Grounding to Grounded Theory: Toward Multi-Grounded Theory". *International Journal of Qualitative Methods*, *17*(1). https://doi.org/10.1177/1609406918795540

Goldkuhl, G., & Karlsson, F. (2020). Method Engineering as Design Science. *Journal of the Association for Information Systems*, *21*(5). https://doi.org/10.17705/1jais.00636

Goldkuhl, G., & Sjöström, J. (2021). Design Science Theorizing: The Contribution of Practical Theory. In N. R. Hassan & L. P. Willcocks (Eds.), *Advancing information systems theories: Rationale and processes* (pp. 239–273). Springer International Publishing. https://doi.org/10.1007/978-3-030-64884-8_7

Goodwin, P., & Wright, G. (2010). The limits of forecasting methods in anticipating rare events. *Technological Forecasting and Social Change*, *77*(3), 355–368. https://doi.org/10.1016/j.techfore.2009.10.008

Graham, S. (2011). Disrupted Cities: Infrastructure Disruptions as the Achilles Heel of Urbanized Societies, 15.

Gregor, S. (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, *30*(3), 611–642.

Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). The Anatomy of a Design Principle. *Journal of the Association for Information Systems*.

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *32*(2), 337–355.

Gregor, S., & Jones, D. (2007). The Anatomy of a Design Theory. *Journal of the Association for Information Systems*, *8*(5), 312–335.

Gregory, S. A. (1966). Design Science. In S. A. Gregory (Ed.), *The design method* (pp. 323–330). Springer US. https://doi.org/10.1007/978-1-4899-6331-4_35

Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M. C., Shyamsundar, P., Steffen, W., Glaser, G., Kanie, N., & Noble, I. (2013). Sustainable development goals for people and planet. *Nature*, *495*, 305.

Grunwald, A. (2007). Converging technologies: Visions, increased contingencies of the conditio humana, and search for orientation. *Futures*, *39*(4), 380–392. https://doi.org/10.1016/j.futures.2006.08.001

Harari, Y., Harris, T., & Raskin, A. (2023). Opinion — You Can Have the Blue Pill or the Red Pill, and We're Out of Blue Pills [newspaper]. *The New York Times*. Retrieved April 16, 2023, from https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html

Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. *Electronic Markets*, *30*(1), 87–98. https://doi.org/10.1007/s12525-019-00377-4

Herwix, A., & Haj-Bolouri, A. (2020). Having a Positive Impact with Design Science Research – Learning from Effective Altruism. In S. Hofmann, O. Müller, & M. Rossi (Eds.), *Designing for digital transformation. co-creating services with citizens and industry* (pp. 235–246). Springer International Publishing. https://doi.org/10.1007/978-3-030-64823-7_22

Herwix, A., & Haj-Bolouri, A. (2021). Revisiting the Problem of the Problem – An Ontology and Framework for Problem Assessment in IS Research. *Proceedings of the Twenty-Nineth European Conference on Information Systems (ECIS2021)*.

Herwix, A., & Rosenkranz, C. (2018). Making Sense of Design Science in Information Systems Research: Insights from a Systematic Literature Review. In S. Chatterjee, K. Dutta, & R. P. Sundarraj (Eds.), *Designing for a digital and globalized world* (pp. 51–66). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-91800-6_4

Herwix, A., Rossi, M., Purao, S., Haj-Bolouri, A., Tremblay, M. C., & Gregor, S. (2022). Ethics in Information Systems and Design Science Research: Five Perspectives. *Communications of the Association for Information Systems*, *50*(1), 589–616. https://doi.org/10.17705/1CAIS.05028

Herwix, A., & Zur Heiden, P. (2022). Context in Design Science Research: Taxonomy and Framework. *Proceedings of the 55th Annual Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2022.705

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science In Information Systems Research. *MIS Quarterly*, *28*(1), 75–105.

Hildebrand, D. (2023). John Dewey. In E. N. Zalta & U. Nodelman (Eds.), *The stanford encyclopedia of philosophy* (Fall 2023). Metaphysics Research Lab, Stanford University. Retrieved September 23, 2023, from https://plato.stanford.edu/archives/fall2023/entries/dewey/

Hillis, D. (2010). The Age of Digital Entanglement. *Scientific American*, *303*(3), 93–93.

Hovorka, D. S. (2009). Design Science Research: A call for a pragmatic perspective. *Proceedings of the Second Scientific Meeting of AIS SIGPrag*.

Hubka, V., & Eder, W. E. (1987). A scientific approach to engineering design. *Design Studies*, *8*(3), 123–137. https://doi.org/10.1016/0142-694X(87)90035-4

Huff, A. G., Beyeler, W. E., Kelley, N. S., & McNitt, J. A. (2015). How resilient is the United States' food system to pandemics? *Journal of Environmental Studies and Sciences*, *5*(3), 337–347. https://doi.org/10.1007/s13412-015-0275-3

IC3. (2021). *Internet Crime Report 2020*. FBI. Retrieved May 28, 2021, from https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf

Iivari, J. (1991). A paradigmatic analysis of contemporary schools of IS development. *European Journal of Information Systems*, *1*(4), 249–272. https://doi.org/10.1057/ejis.1991.47

Iivari, J. (2007). A Paradigmatic Analysis of Information Systems As a Design Science. *Scandinavian Journal of Information Systems*, *19*(2). http://aisel.aisnet.org/sjis/vol19/iss2/5

Iivari, J., Hirschheim, R., & Klein, H. K. (1998). A Paradigmatic Analysis Contrasting Information Systems Development Approaches and Methodologies. *Information Systems Research*, *9*(2), 164–193. https://doi.org/10.1287/isre.9.2.164

Janardhan, S. (2021, October 5). More details about the October 4 outage. Retrieved November 3, 2023, from https://engineering.fb.com/2021/10/05/networking-traffic/outage-details/

Jirotka, M., Grimpe, B., Stahl, B. C., Eden, G., & Hartswood, M. (2017). Responsible research and innovation in the digital age. *Communications of the ACM*, *60*(5), 62–68. https://doi.org/10.1145/3064940

Jonas, H. (1984). *The imperative of responsibility: in search of an ethics for the technological age*. University of Chicago Press.

Keen, P. G. (1991). Relevance and rigor in information systems research: Improving quality, confidence, cohesion and impact. In H.-E. Nissen, H. K. Klein, & R. Hirschheim (Eds.), *Information systems research: Contemporary approaches and emergent traditions* (p. 49).

King, J. L., & Kraemer, K. L. (2019). Policy: An Information Systems Frontier. *Journal of the Association for Information Systems*, *20*(6), 842–847. https://doi.org/10.17705/1jais.00553

Kroeze, J. H., Travica, B., & Van Zyl, I. (2019). Information Systems in a Transdisciplinary Perspective: Leaping to a Larger Stage. *Alternation: Interdisciplinary Journal for the Study of the Arts and Humanities in Southern Africa*, *Sp24*. https://doi.org/10.29086/2519-5476/2019/sp24.2a2

Kuechler, B., & Vaishnavi, V. (2008). On theory development in design science research: Anatomy of a research project. *European Journal of Information Systems*, *17*(5), 489–504. https://doi.org/10.1057/ejis.2008.40

Kuechler, W., Vaishnavi, V. K., & Petter, S. (2005). The Aggregate General Design Cycle as a Perspective on the Evolution of Computing Communities of Interest. *Computing Letters*, *1*(3), 123–128. https://doi.org/10.1163/1574040054861221

Kuechler, W., & Vaishnavi, V. (2008). The emergence of design research in information systems in North America. *Journal of Design Research*, *7*(1), 1–16.

Kuechler, W., & Vaishnavi, V. (2012). A Framework for Theory Development in Design Science Research: Multiple Perspectives. *Journal of the Association for Information Systems*, *13*(6), 395–423.

Kuhn, T. S. (2012, April 18). *The Structure of Scientific Revolutions: 50th Anniversary Edition*. University of Chicago Press.

Lee, A. S. (2000). *Systems Thinking, Design Science, and Paradigms: Heeding Three Lessons from the Past to Resolve Three Dilemmas in the Present to Direct a Trajectory for Future Research in the Information Systems Field*. Retrieved February 22, 2022, from http://www.people.vcu.edu/~aslee/ICIM-keynote-2000/ICIM-keynote-2000.htm

Lee, A. S., & Nickerson, J. V. (2010). Theory As a Case of Design: Lessons for Design from the Philosophy of Science. *Proceedings of the 43rd Hawaii International Conference on System Sciences*, 1–8. https://doi.org/10.1109/HICSS.2010.484

Lee, A. S., Thomas, M., & Baskerville, R. L. (2015). Going back to basics in design science: From the information technology artifact to the information systems artifact. *Information Systems Journal*, *25*(1), 5–21. https://doi.org/10.1111/isj.12054

Levin, K., Cashore, B., Bernstein, S., & Auld, G. (2012). Overcoming the tragedy of super wicked problems: Constraining our future selves to ameliorate global climate change. *Policy Sciences*, *45*(2), 123–152. https://doi.org/10.1007/s11077-012-9151-0

Levy, M., & Hirschheim, R. (2012). Removing the Positivist Straight Jacket From Information Systems Design Science Research. *ECIS 2012 Proceedings*.

Lewis, S. L., & Maslin, M. A. (2015). Defining the Anthropocene. *Nature*, *519*(7542), 171–180. https://doi.org/10.1038/nature14258

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2017). Paradigmatic Controversies, Contradictions, and Emerging Confluences, Revisited. *The sage handbook of qualitative research* (pp. 213–263). SAGE Publications.

Lukyanenko, R., & Parsons, J. (2020). Design Theory Indeterminacy: What is it, how can it be reduced, and why did the polar bear drown? *Journal of the Association for Information Systems*, *21*(5), 62.

Maedche, A. (2017). Interview with Prof. Jeroen van den Hoven on "Why do Ethics and Values Matter in Business and Information Systems Engineering?". *Business & Information Systems Engineering*, *59*(4), 297–300. https://doi.org/10.1007/s12599-017-0476-2

Mahmoudi, M., & Nadi, S. (2018). The Android update problem: An empirical study. *Proceedings of the 15th International Conference on Mining Software Repositories*, 220–230. https://doi.org/10.1145/3196398.3196434

Majchrzak, A., Markus, M. L., & Wareham, J. (2016). Designing for Digital Transformation: Lessons for Information Systems Research from the Study of ICT and Societal Challenges. *MIS Quarterly*, *40*(2), 267–277. https://doi.org/10.25300/MISQ/2016/40:2.03

Malhotra, A., Melville, N. P., & Watson, R. T. (2013). Spurring Impactful Research on Information Systems and Environmental Sustainability. *MIS Quarterly*, *37*(4), 1265–1274. https://doi.org/10.25300/MISQ/2013/37:4.3

Manheim, D. (2020). The Fragile World Hypothesis: Complexity, Fragility, and Systemic Existential Risk. *Futures*, *122*, 102570. https://doi.org/10.1016/j.futures.2020.102570

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*, 251–266.

McKay, J., Marshall, P., & Hirschheim, R. (2012). The design construct in information systems design science. *Journal of Information Technology*, *27*(2), 125–139.

Metcalfe, B. (2013). Metcalfe's Law after 40 Years of Ethernet. *Computer*, *46*(12), 26–31. https://doi.org/10.1109/MC.2013.374

Meyer, B. (2018). Making Sense of Agile Methods. *IEEE Software*, *35*(2), 91–94. https://doi.org/10.1109/MS.2018.1661325

Moeini, M., Rahrovani, Y., & Chan, Y. E. (2019). A review of the practical relevance of IS strategy scholarly research. *The Journal of Strategic Information Systems*, *28*(2), 196–217. https://doi.org/10.1016/j.jsis.2018.12.003

Moor, J. H. (1985). What Is Computer Ethics? *Metaphilosophy*, *16*(4), 266–275. https://doi.org/10.1111/j.1467-9973.1985.tb00173.x

Morana, S., vom Brocke, J., Maedche, A., Seidel, S., Adam, M. T. P., Bub, U., Fettke, P., Gau, M., Herwix, A., Mullarkey, M. T., Nguyen, H. D., Sjöström, J., Toreini, P., Wessel, L., & Winter, R. (2018). Tool Support for Design Science Research—Towards a Software Ecosystem: A Report from a DESRIST 2017 Workshop. *Communications of the Association for Information Systems*, *43*(1), 237–256. https://doi.org/10.17705/1CAIS.04317

Morgan, D. L. (2007). Paradigms Lost and Pragmatism Regained. *Journal of Mixed Methods Research*, *1*(1), 48–76.

Morgan, D. L. (2014). Pragmatism as a Paradigm for Social Research. *Qualitative Inquiry*, *20*(8), 1045–1053. https://doi.org/10.1177/1077800413513733

Mumford, E. (1995). *Effective systems design and requirements analysis: The ETHICS approach*. Macmillan Education UK.

Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science research in information systems. *Information & Management*, *51*(6), 801–809. https://doi.org/10.1016/j.im.2014.01.002

Niehaves, B., & Ortbach, K. (2016). The inner and the outer model in explanatory design theory: The case of designing electronic feedback systems. *European Journal of Information Systems*, *25*(4), 303–316. https://doi.org/10.1057/ejis.2016.3

Novitzky, P., Bernstein, M. J., Blok, V., Braun, R., Chan, T. T., Lamers, W., Loeber, A., Meijer, I., Lindner, R., & Griessler, E. (2020). Improve alignment of research policy and societal values. *Science*, *369*(6499), 39–41. https://doi.org/10.1126/science.abb3415

Nunamaker, J. F., & Briggs, R. O. (2011). Toward a broader vision for information systems. *ACM Transactions on Management Information Systems (TMIS)*, *2*(4), 20. https://doi.org/10.1145/2070710.2070711

Nunamaker, J. F., Briggs, R. O., Derrick, D. C., & Schwabe, G. (2015). The Last Research Mile: Achieving Both Rigor and Relevance in Information Systems Research. *Journal of Management Information Systems*, *32*(3), 10–47.

Nunamaker, J. F., Chen, M., & Purdin, T. D. (1990). Systems Development in Information Systems Research. *Journal of Management Information Systems*, *7*(3), 89–106.

Nunamaker, J. F., Twyman, N. W., & Giboney, J. S. (2013). Breaking out of the Design Science Box: High-Value Impact Through Multidisciplinary Design Science Programs of Research. *AMCIS 2013 Proceedings*, 11.

Nunamaker, J. F., Twyman, N. W., Giboney, J. S., & Briggs, R. O. (2017). Creating High-Value Real-World Impact Through Systematic Programs of Research. *MIS Quarterly*, *41*(2), 335–351.

OpenAI. (2023, March 27). Gpt-4 Technical Report. https://doi.org/10.48550/arXiv.2303.08774

Ord, T. (2020). *The precipice: Existential risk and the future of humanity*. Hachette Books.

Orlikowski, W. J., & Iacono, C. S. (2001). Research Commentary: Desperately Seeking the "IT" in IT Research—A Call to Theorizing the IT Artifact. *Information Systems Research*, *12*(2), 121–134. https://doi.org/10.1287/isre.12.2.121.9700

Ormerod, R. J. (2006). The history and ideas of pragmatism. *Journal of the Operational Research Society*, *57*(8), 892–909. https://doi.org/10.1057/palgrave.jors.2602065

Ormerod, R. J. (2020). Pragmatism in professional practice. *Systems Research and Behavioral Science*, *n/a*(n/a). https://doi.org/10.1002/sres.2739

Owen, R. (2019). Responsible Innovation and Responsible Research and Innovation. *Handbook on science and public policy* (pp. 26–48). Edward Elgar Publishing. https://doi.org/10.4337/9781784715946.00010

Owen, R., Bessant, J., & Heintz, M. (Eds.). (2013). *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society* (1st ed.). Wiley.

Owen, R., Pansera, M., Macnaghten, P., & Randles, S. (2021). Organisational institutionalisation of responsible innovation. *Research Policy*, *50*(1), 104132. https://doi.org/10.1016/j.respol.2020.104132

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013, April 2). A Framework for Responsible Innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible innovation: Managing the responsible emergence of science and innovation in society* (pp. 27–50). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118551424.ch2

Paradice, D., Parrish, J., & Richardson, S. (2019). Grand Challenge Pursuits: Insights from a Multi-year DSR Project Stream. *Communications of the Association for Information Systems*, *45*.

Parks-Leduc, L., Mulligan, L., & Rutherford, M. A. (2021). Can Ethics Be Taught? Examining the Impact of Distributed Ethical Training and Individual Characteristics on Ethical Decision-Making. *Academy of Management Learning & Education*, *20*(1), 30–49. https://doi.org/10.5465/amle.2018.0157

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77.

Perrow, C. (2011, October 12). *Normal Accidents: Living with High Risk Technologies*. Princeton University Press.

Poel, I. v. d., Royakkers, L., & Zwart, S. D. (2015). *Moral responsibility and the problem of many hands* (1 [edition]). Routledge.

Purao, S., Baldwin, C., Hevner, A. R., Storey, V. C., Pries-Heje, J., Smith, B., & Zhu, Y. (2008). The Sciences of Design: Observations on an Emerging Field. *Communications of the Association for Information Systems*, *23*, 523–546.

Rai, A. (2017a). Diversity of Design Science Research. *MIS Quarterly*, *41*(1), iii–xviii.

Rai, A. (2017b). Editor's Comments: Avoiding Type III Errors: Formulating IS Research Problems that Matter. *MIS Quarterly*, *41*(2), iii–vii.

Ram, S., & Goes, P. B. (2021). Focusing on Programmatic High Impact Information Systems Research, Not Theory, to Address Grand Challenges. *MIS Quarterly*, *45*(1), 479–483.

Randles, S., Larédo, P., Loconto, A. M., Walhout, B., & Lindner, R. (2016). Framings and frameworks: Six grand narratives of de facto RRI. *Navigating towards shared responsibility in research and innovation. approach, process and results of the res-agora project* (np). Fraunhofer Institute for Systems; Innovation Research (ISI). Retrieved September 7, 2023, from https://hal.science/hal-01320462

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, *4*(2), 155–169.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S. I., Lambin, E., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., De Wit, C. A., Hughes, T., Van Der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., . . . Foley, J. (2009). Planetary Boundaries: Exploring the Safe Operating Space for Humanity. *Ecology and Society*, *14*(2), art32. https://doi.org/10.5751/ES-03180-140232

Rohde, M., Broedner, P., Stevens, G., Betz, M., & Wulf, V. (2017). Grounded design - a praxeological is research perspective. *JOURNAL OF INFORMATION TECHNOLOGY*, *32*(2), 163–179. https://doi.org/10.1057/jit.2016.5

Ross, C., & Swetlitz, I. (2017). Ibm pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close [magazine]. *STAT*. Retrieved April 12, 2021, from https://www.statnews.com/2017/09/05/watson-ibm-cancer/

Ross-Hellauer, T., Deppe, A., & Schmidt, B. (2017). Survey on open peer review: Attitudes and experience amongst editors, authors and reviewers. *PLOS ONE*, *12*(12), e0189311. https://doi.org/10.1371/journal.pone.0189311

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Sanger, D. E., & Perlroth, N. (2021). Pipeline Attack Yields Urgent Lessons About U.S. Cybersecurity [newspaper]. *The New York Times*. Retrieved May 28, 2021, from https://www.nytimes.com/2021/05/14/us/politics/pipeline-hack.html

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and Its Continued Relevance. *MIS Quarterly*, *43*(3), 695–A5. https://doi.org/10.25300/MISQ/2019/13747

Satariano, A. (2021). What is Fastly, the company behind the worldwide internet outage? [newspaper]. *The New York Times*. Retrieved June 15, 2021, from https://www.nytimes.com/2021/06/08/business/fastly-internet-outage.html

Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., Koppel, J. v. d., Leemput, I. A. v. d., Levin, S. A., Nes, E. H. v., Pascual, M., & Vandermeer, J. (2012). Anticipating Critical Transitions. *Science*, *338*(6105), 344–348. https://doi.org/10.1126/science.1225244

Schroeder, D., & Kaplan, D. (2019). Responsible Inclusive Innovation: Tackling Grand Challenges Globally. In R. von Schomberg & J. Hankins (Eds.), *International handbook on responsible innovation: A global resource* (pp. 308–324). Edward Elgar Publishing.

Seger, E., Avin, S., Pearson, G., Briers, M., Ó Heigeartaigh, S., & Bacon, H. (2020). *Tackling threats to informed decision-making in democratic societies – Promoting epistemic security in a technologically-advanced world*. Alan Turing Institute. Retrieved November 4, 2020, from https://www.turing.ac.uk/sites/default/files/2020-10/epistemic-security-report_final.pdf

Seidel, S., & Watson, R. (2020). Integrating Explanatory/Predictive and Prescriptive Science in Information Systems Research. *Communications of the Association for Information Systems*, *47*(1). https://doi.org/10.17705/1CAIS.04714

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, *35*, 37–56.

Shanley, D., Cohen, J. B., Surber, N., & Stack, S. (2022). Looking beyond the 'horizon' of RRI: moving from discomforts to commitments as early career researchers. *Journal of Responsible Innovation*, *9*(1), 124–132. https://doi.org/10.1080/23299460.2022.2049506

Shilton, K., Finn, M., & DuPont, Q. (2021). Shaping ethical computing cultures. *Communications of the ACM*, *64*(11), 26–29. https://doi.org/10.1145/3486639

Simon, H. A. (1973). The Structure of Ill Structured Problems. *Artificial Intelligence*, 21.

Simon, H. A. (1996). *The Sciences of the Artificial* (3rd). MIT Press.

Sjöström, J. (2010). *Designing Information Systems: A pragmatic account* (Doctoral dissertation). Uppsala, Sweden, Uppsala University. http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-130661

Stahl, B. C. (2012). Responsible research and innovation in information systems. *European Journal of Information Systems*, *21*(3), 207–211. https://doi.org/10.1057/ejis.2012.19

Stahl, B. C. (2022). Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. *International Journal of Information Management*, *62*, 102441. https://doi.org/10.1016/j.ijinfomgt.2021.102441

Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387–420. https://doi.org/10.1177/030631289019003001

Steen, M., Timan, T., & van de Poel, I. (2021). Responsible innovation, anticipation and responsiveness: Case studies of algorithms in decision support in justice and security, and an exploration of potential, unintended, undesirable, higher-order effects. *AI and Ethics*, *1*(4), 501–515. https://doi.org/10.1007/s43681-021-00063-2

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Stockton, P., & Council, E. (2016). *Electric Infrastructure Protection (E-PRO®) Handbook* (Vol. 1). Electric Infrastructure Security (EIS) Council.

Stockton, P., & Council, E. (2018). *Electric Infrastructure Protection (E-PRO®) Handbook III* (Vol. 3). Electric Infrastructure Security (EIS) Council.

Strickland, E. (2019). Ibm Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, *56*(4), 24–31. https://doi.org/10.1109/MSPEC.2019.8678513

Sun, R., & Gregor, S. (2023). Reconceptualizing platforms in information systems research through the lens of service-dominant logic. *The Journal of Strategic Information Systems*, *32*(3), 101791. https://doi.org/10.1016/j.jsis.2023.101791

Tabarés, R., Loeber, A., Nieminen, M., Bernstein, M. J., Griessler, E., Blok, V., Cohen, J., Hönigmayer, H., Wunderle, U., & Frankus, E. (2022). Challenges in the implementation of responsible research and innovation across Horizon 2020. *Journal of Responsible Innovation*, *9*(3), 291–314. https://doi.org/10.1080/23299460.2022.2101211

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder*. Random House.

Tarafdar, M., Gupta, A., & Turel, O. (2013). The dark side of information technology use. *Information Systems Journal*, *23*(3), 269–275. https://doi.org/10.1111/isj.12015

Tarafdar, M., Shan, G., Bennett Thatcher, J., & Gupta, A. (2022). Intellectual Diversity in IS Research: Discipline-Based Conceptualization and an Illustration from Information Systems Research. *Information Systems Research*, *33*(4), 1490–1510. https://doi.org/10.1287/isre.2022.1176

Thiel, D. V. (Ed.). (2014). Introduction to engineering research. In *Research methods for engineers* (pp. 1–26). Cambridge University Press. https://doi.org/10.1017/CBO9781139542326.003

Timmermans, J., & Blok, V. (2021). A critical hermeneutic reflection on the paradigm-level assumptions underlying responsible innovation. *Synthese*, *198*(19), 4635–4666. https://doi.org/10.1007/s11229-018-1839-z

Timmermans, J., Blok, V., Braun, R., Wesselink, R., & Nielsen, R. Ø. (2020). Social labs as an inclusive methodology to implement and study social change: The case of responsible research and innovation. *Journal of Responsible Innovation*, *7*(3), 410–426. https://doi.org/10.1080/23299460.2020.1787751

Ulrich, W. (2003). Beyond methodology choice: Critical systems thinking as critically systemic discourse. *Journal of the Operational Research Society*, *54*(4), 325–342. https://doi.org/10.1057/palgrave.jors.2601518

Ulrich, W. (1994). Can We Secure Future-Responsive Management Through Systems Thinking and Design? *Interfaces*, *24*(4), 26–37. https://doi.org/10.1287/inte.24.4.26

Ulrich, W. (2006). Critical Pragmatism: A New Approach to Professional and Business Ethics. *Interdisciplinary yearbook for business ethics. v. 1, v. 1*. Peter Lang Pub Inc.

Ulrich, W. (2016). Philosophy for professionals: Towards critical pragmatism [magazine]. *Ulrich's Bimonthly*, *14*(2), 1–17.

Ulrich, W., & Reynolds, M. (2010). Critical Systems Heuristics. In M. Reynolds & S. Holwell (Eds.), *Systems approaches to managing change: A practical guide* (pp. 243–292). Springer London. https://doi.org/10.1007/978-1-84882-809-4_6

Vaishnavi, V., Kuechler, B., & Petter, S. (2019). *Design Science Research in Information Systems*. Retrieved October 28, 2020, from http://www.desrist.org/desrist/content/design-science-research-in-information-systems.pdf

Vaishnavi, V., & Kuechler, W. (2005, February 8). *Design Research in Information Systems*. Retrieved August 26, 2023, from https://web.archive.org/web/20050208031622/http://www.isworld.org/Researchdesign/drisISworld.htm

Vaishnavi, V., & Kuechler, W. J. (2015). *Design Science Research Methods and Patterns* (2nd Edition). CRC Press.

Venable, J. (2006). The role of theory and theorising in design science research. *Proceedings of the 1st International Conference on Design Science in Information Systems and Technology*, 1–18.

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*(1), 233. https://doi.org/10.1038/s41467-019-14108-y

Voegtlin, C., & Scherer, A. G. (2017). Responsible Innovation and the Innovation of Responsibility: Governing Sustainable Development in a Globalized World. *Journal of Business Ethics*, *143*(2), 227–243. https://doi.org/10.1007/s10551-015-2769-z

vom Brocke, J., Winter, R., Hevner, A. R., & Maedche, A. (2020). Accumulation and Evolution of Design Knowledge in Design Science Research – A Journey Through Time and Space. *Journal of the Association for Information Systems*, *21*(3), 520–544. https://doi.org/10.17705/1jais.00611

von Schomberg, R. (2013, May 20). A Vision of Responsible Research and Innovation. *Responsible innovation: Managing the responsible emergence of science and innovation in society*. Wiley. Retrieved January 6, 2020, from https://papers.ssrn.com/abstract=2428157

von Schomberg, R. (2019). Why responsible innovation? In R. von Schomberg & J. Hankins (Eds.), *International handbook on responsible innovation: A global resource* (pp. 12–32). Edward Elgar Publishing Cheltenham, UK.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (2004). Assessing information system design theory in perspective: How useful was our 1992 initial rendition? *JITTA: Journal of Information Technology Theory and Application*, *6*(2).

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research*, *3*(1), 36–59.

Walsham, G. (2012). Are we making a better world with ICTs? Reflections on a future agenda for the IS field. *Journal of Information Technology*, *27*(2), 87–93. https://doi.org/10.1057/jit.2012.4

Walsham, G. (2017). Ict4d research: Reflections on history and future agenda. *Information Technology for Development*, *23*(1), 18–41. https://doi.org/10.1080/02681102.2016.1246406

Wang, P. (2021). Connecting the Parts with the Whole: Toward an Information Ecology Theory of Digital Innovation Ecosystems. *MIS Quarterly*, *45*(1), 397–422.

Wanzenböck, I., Wesseling, J. H., Frenken, K., Hekkert, M. P., & Weber, K. M. (2020). A framework for mission-oriented innovation policy: Alternative pathways through the problem–solution space. *Science and Public Policy*, scaa027. https://doi.org/10.1093/scipol/scaa027

Watson, D., & Floridi, L. (2018). Crowdsourced science: Sociotechnical epistemology in the e-research paradigm. *Synthese*, *195*(2), 741–764. https://doi.org/10.1007/s11229-016-1238-2

Weber, R. (1987). Toward A Theory of Artifacts: A Paradigmatic Base For Information Systems Research. *Journal of Information Systems*, *1*(2), 3.

Wessel, L., Peters, C., & Niederman, F. (2021). Using 'Panel Reports' to Advance Scholarly Discourse: A Change in Editorial Policy and Guidelines for Authors of 'Panel Reports'. *Communications of the AIS*. https://doi.org/10.17705/1CAIS.04915

Wilkinson, A., Kupers, R., & Mangalagiu, D. (2013). How plausibility-based scenario practices are grappling with complexity to appreciate and address 21st century challenges. *Technological Forecasting and Social Change*, *80*(4), 699–710. https://doi.org/10.1016/j.techfore.2012.10.031

Wilson, D. S., Madhavan, G., Gelfand, M. J., Hayes, S. C., Atkins, P. W. B., & Colwell, R. R. (2023). Multilevel cultural evolution: From new theory to practical applications. *Proceedings of the National Academy of Sciences*, *120*(16), e2218222120. https://doi.org/10.1073/pnas.2218222120

Wilson, E. O. (1999, March 30). *Consilience: The Unity of Knowledge* (Reprint edition). Vintage.

Winter, S. J., & Butler, B. S. (2011). Creating Bigger Problems: Grand Challenges as Boundary Objects and the Legitimacy of the Information Systems Field. *Journal of Information Technology*, *26*(2), 99–108. https://doi.org/10.1057/jit.2011.6

With great power comes great responsibility. (2023, June 14). In *Wikipedia*. Retrieved June 17, 2023, from https://en.wikipedia.org/w/index.php?title=With_great_power_comes_great_responsibility&oldid=1160041089

Wolff, J. (2020). Opinion — How Ransomware Puts Your Hospital at Risk [newspaper]. *The New York Times*. Retrieved May 28, 2021, from https://www.nytimes.com/2020/10/17/opinion/hospital-internet-security-ransomware.html

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting*, *25*(4), 813–825. https://doi.org/10.1016/j.ijforecast.2009.05.019

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research commentary—the new organizing logic of digital innovation: An agenda for information systems research. *Information systems research*, *21*(4), 724–735.

You, X., & Hands, D. (2019). A Reflection upon Herbert Simon's Vision of Design in The Sciences of the Artificial. *The Design Journal*, *22*(sup1), 1345–1356. https://doi.org/10.1080/14606925.2019.1594961

Yudkowsky, E. (2008, July 3). Artificial Intelligence as a positive and negative factor in global risk. In N. Bostrom & M. M. Cirkovic (Eds.), *Global catastrophic risks*. Oxford University Press. https://doi.org/10.1093/oso/9780198570509.003.0021

Yudkowsky, E. (2022). Agi Ruin: A List of Lethalities. Retrieved August 25, 2023, from https://www.alignmentforum.org/posts/uMQ3cqWDPHhjtiesc/agi-ruin-a-list-of-lethalities

Publications

Ethics in Information Systems and Design Science Research: Five Perspectives

Outline

Abstract

While ethics are recognized as an integral part of information systems (IS) research, many questions about the role of ethics in research practice remain unanswered. Our report responds to this emerging set of concerns with a broad and integrative account of five perspectives on ethics in IS research and design science research (DSR) in particular. Our report is informed by a broad literature review, a panel discussion at DESRIST 2020, and substantial personal experience from wrestling with ethical considerations in the field. The report provides a comprehensive discussion of prevailing perspectives on ethics and draws implications for IS research. Together, we hope the report will inspire more ethics-conscious and responsible IS research.

Keywords: Ethics, Information Systems Field, Design Science Research, Panel Report.

Bibliographic Information

Herwix, A., Rossi, M., Purao, S., Haj-Bolouri, A., Tremblay, M. C., & Gregor, S. (2022). Ethics in Information Systems and Design Science Research: Five Perspectives. *Communications of the Association for Information Systems*, 50(1), 589–616. https://doi.org/10.17705/1CAIS.05028

Author's contribution

Conceptualization (lead), Methodology (lead), Writing – original draft (lead), Writing – review and editing (lead), Visualization (lead), Project administration (lead).

Copyright Notice

1 Introduction

Ongoing digital transformation (Vial, 2019) and emerging technological capabilities have far reaching implications for all facets of societies, and therefore, raise ethical concerns about how we *should* deal with such developments (European Group on Ethics in Science and New Technologies, 2018). As van den Hoven describes in an interview about the role of ethics in information systems (IS):

> Every design, artifact, system is shaped by the values, ideas and world views of the designer and builder. That applies to architecture, software engineering, product design, synthetic biology, material science and civil engineering. A design is a consolidated set of choices made by designers, developers and engineers. Via their designs for systems and artifacts they come to have an incredible impact on the lives of others [. . .]. These are all formidable shapers of the world we inhabit and in which we acquire our beliefs, decide and act, expect, feel, and hope. [. . .] In order to shape these environments in which we will function as moral beings in a responsible way, we need to express or ''design in'' our shared moral values. Values should therefore be seen as a sort of supra- or non-functional requirements for which we can and ought to design. It will become more and more important in the future to be able to design systematically for moral, legal and social requirements. (Maedche, 2017, p. 298)

Given this overt importance of values and ethics for understanding, managing, and designing digital systems, it is clear that ethics—the study of what is right and wrong ("ETHICS", n.d.) in terms of "a set of personal or social standards for good or bad behavior and character" ("MORALITY", n.d.)—should be an important consideration for *all* IS scholars (Mingers & Walsham, 2010; Paradice et al., 2018; Porra, 2001; Stahl, 2012a; Walsham, 1996, 2012). Moreover, ethical considerations and questions are especially *prominent* for scholars engaged in design science research (DSR) because they aim to change the world by solving real world problems (Hevner et al., 2004; Purao, 2013; Walsham, 2012). In DSR, scholars are explicitly engaged in ethical questions throughout the research lifecycle (e.g., Benke et al., 2020; Herwix & Haj-Bolouri, 2020; Myers & Venable, 2014; Stahl et al., 2011). For instance, which problems should we prioritize given constraints on available resources (e.g., Herwix & Haj-Bolouri, 2020, 2021; Purao, 2021)? How can a DSR project be carried out in an ethical manner (e.g., Benke et al., 2020; Myers & Venable, 2014)? What are appropriate techniques for integrating ethical considerations into the design of solutions (e.g., Friedman et al., 2008; Manders-Huits, 2011a; van den Hoven, 2007)? How can evaluations be leveraged to enhance the ethical value of research projects (e.g., Stahl et al., 2011)? Questions such as these are critical for design science researchers, and, in our opinion, rarely addressed.

With this report, we wish to raise awareness of these concerns, and contribute to the nascent stream of research on engaging with ethics in IS research. Our effort in this report is to develop a comprehensive discussion of ethical considerations for IS research. In particular, our report complements studies on specific aspects or dimensions of ethics in IS research with an overarching discussion informed by five intersecting perspectives (see Figure 7):

1. the *Philosophy* perspective concerned with the fundamental questions at the core of ethics,
2. the *Design* perspective concerned with the ethics of design and technology,

3. the *Science* perspective concerned with ethics in the realm of science,

4. the *IS Practice* perspective concerned with ethics as applied in IS practice, and

5. the *Panelists* perspectives concerned with the unique challenges of ethics in DSR practice.

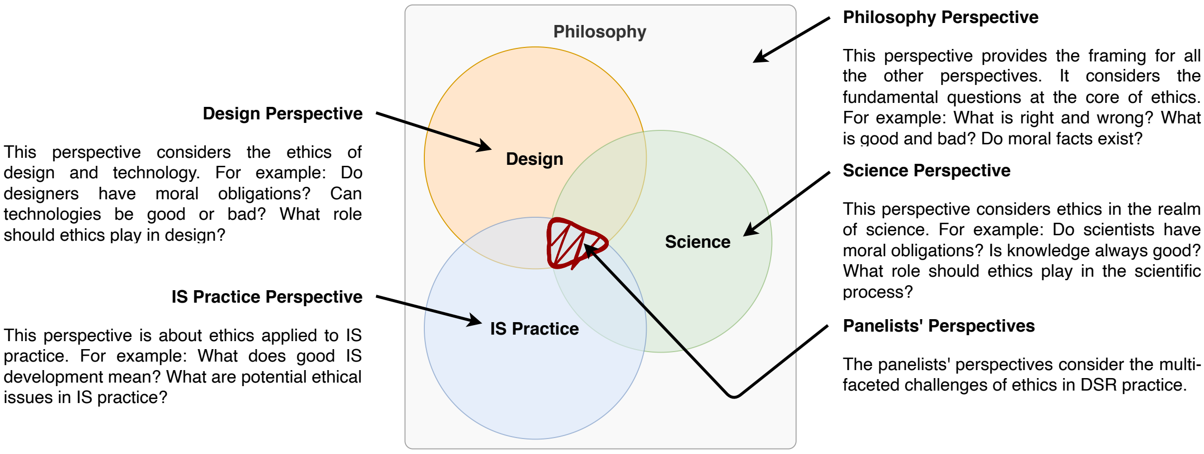

Figure 7: Five Perspectives on Ethics in IS and DSR.

Our exploration of these perspectives is inspired by a panel discussion at the International Conference on Design Science Research in Information Systems and Technology (DESRIST 2020) which brought together four eminent DSR scholars: Matti Rossi, Monica Chiarini Tremblay, Sandeep Purao, and Shirley Gregor. Following the panel discussion with a focus on DSR, we conducted a broad review of ethics that elaborated on and substantiated the concerns discussed by the panelists as well as expanded their scope beyond DSR. The resulting report should be of interest to all IS researchers and not just those interested in DSR.

The rest of the paper is organized as follows. Section 2 presents a broad overview of ethics in philosophy along the three major areas of research: metaethics, normative ethics, and applied ethics. In section 3, we discuss the role of ethics in design exploring the themes of ethics-conscious design, technological risks and unintended consequences, and responsibility in technology design. Section 4 reviews some major developments of ethics in science. In section 5, we derive and discuss challenges for ethics in IS practice based on extant literature and the authors professional experience. The panelists' personal views on ethics are presented in section 6. In Section 7 we integrate the discussion of the preceding perspectives and synthesize a set of recommendations to advance the IS field. We conclude in section 8 by offering a call for action.

2 Ethics in Philosophy

As Singer (2019) explains, ethics is a branch of philosophy focused on understanding and justifying concepts of right and wrong as well as good and bad behavior. It is closely linked to the branch of *aesthetics*—the study of beauty and taste—as both are concerned with *axiology*—the study of values. The question of how to behave is fundamental to human life in social groups. Thus, it is not surprising that the origin of ethics can be traced back to ancient times with many ethical theories and perspectives evolving over time. Today, the study of ethics is grouped into three major areas of research:

- *Metaethics*, focused on studying the nature, scope, and meaning of moral judgment,
- *Normative ethics*, focused on studying how one ought to be and act in general, and
- *Applied ethics*, focused on studying how to act in a specific situation and circumstance.

Discussing these streams of research in depth is well beyond the scope of this paper. In the following, we provide an overview of important considerations that scholars have discussed in each. For a more comprehensive introduction the reader is referred to the excellent summary of the field of ethics by Singer (2019). A more targeted reflection of how ethics may be related to IS research in particular can be found in Stahl (2012a).

2.1 Metaethics

Metaethics is the branch of ethics concerned with questions about the nature, scope, and meaning of moral judgment. Here, the issues and questions that scholars work on are abstract and a step removed from substantive debates. Metaethics investigates the assumptions and commitments that are reflected in such debates. Thus, it is generally thought to provide a neutral background against which competing moral views can be examined and evaluated (Sayre-McCord, 2014). In the following, we summarize key metaethical positions regarding the *semantics* of moral discourse, the *ontology* of moral properties, the *psychology* of how morality affects us as embodied human agents, and the *epistemology* of how we come to know moral values (DeLapp, n.d.). Understanding and locating oneself in relation to these positions is an important prerequisite for serious and rigorous engagement with ethics. For instance, metaethical positions may have strong implications for how to approach ethical discourse. More elaborate introductions to metaethics can be found in Sayre-McCord (2014) and DeLapp (n.d.).

Semantic Issues in Metaethics: Cognitivism vs. Non-Cognitivism

One of the major semantic issues in metaethics is the question of what exactly people mean when they make moral judgements. Two broad positions can be distinguished (DeLapp, n.d.; Sayre-McCord, 2014):

- *Cognitivism* holds that we are expressing beliefs or making claims that can be either true or false;
- *Non-Cognitivism* holds that we are doing something else (e.g., taking a stand, expressing an emotion, prescribing an action, etc.) which lacks "truth-apt" cognitive content (i.e., it cannot be true or false).

Needless to say, there are interesting arguments for and against both of these positions that go beyond the scope of this paper (e.g., DeLapp, n.d.; Sayre-McCord, 2014). However, we can outline a common challenge that both cognitivism and non-cognitivism face (Sayre-McCord, 2014). Each perspective must account for the observation that moral thought and talk are both (a) distinct (e.g., many people would agree that the moral value of an action cannot be observed in the same way that spatial-temporal objects can) but also (b) significantly continuous with non-moral thought and talk (e.g., moral statements are often phrased in a way that is similar to non-moral "truth-apt" statements). Here and in general, cognitivism more easily accommodates the continuity between moral and non-moral thought and talk but struggles to appreciate the distinctive nature of moral thought and talk. On the other hand, non-cognitivism more easily recognizes the distinctive nature of moral thought and talk but is more challenged to reconcile the continuity.

Ontological Issues in Metaethics: Moral Realism vs. Moral Relativism

Metaethical positions may be distinguished in terms of the ontological status that they ascribe to moral values. Two general positions are often distinguished (DeLapp, n.d.; Sayre-McCord, 2014):

- *Moral realism* (also known as objectivism) holds that moral values are “real” or “objective” in that they exist independently of our beliefs or evidence about them;
- *Moral relativism* (also known as subjectivism) holds that moral values are created by individuals or cultures and are, thus, not belief-independent facts which exist “out there”.

Moral realism is often defended in an effort to establish universal moral truths that exist irrespective of what any individual or culture might think of them. For instance, one might argue if moral truths are not “objective”, does ethics not become arbitrary? However, even within realist positions there is disagreement about what moral values actually are if they are independent from human belief or culture (DeLapp, n.d.).

Moral relativism rejects the idea of moral values as “real” or “objective” in the sense of being independent from belief or culture and maintains that morality is fundamentally anthropocentric. It argues that moral values are created by subjective perspectives and needs on individual or cultural levels and cannot be true or false in a general sense. A relativistic view is often justified based on psychological, epistemological or anthropological considerations (DeLapp, n.d.). For instance, anthropologists have suggested that morality may have emerged from the need for social coordination (Roes, 2003) which is sometimes counted as evidence supporting moral relativism.

We note that the question of moral realism vs. moral relativism is orthogonal to the semantic question of cognitivism vs. non-cognitivism. In particular, it is possible to be a moral relativist and cognitivist at the same time by maintaining that moral language is systematically false (also known as ”error theory”; DeLapp, n.d.).

Psychological Issues in Metaethics: Internalism vs. Externalism

An important psychological issue within metaethics is the question of whether there is an inherent motivation for being moral for its own sake or if other persuasive reasons are needed to act morally. There are two general positions that can be distinguished (DeLapp, n.d.; Sayre-McCord, 2014; Singer, 2019):

- *Internalism* is the position that a moral judgment already implies a psychological motivation to act;
- *Externalism* is the position that an additional desire is needed to motivate moral behavior.

Internalism is represented, for example, in the views of Socrates or Aristotle who held that developing one’s own morality in the form of virtue was at the core of being human. Externalism is often concerned with the apparent conflict or compatibility between morality and self-interest. For instance, while it has been argued that acting morally is generally in the long-term interest of an individual, there may always be a temptation to behave immorally in any particular situation if self-interest and morality are in conflict. A question that is often asked in this context is: “Is it rational to be moral?” Answers diverge and indicate that it may be rational to be, both, immoral and moral given particular assumptions and circumstances.

Epistemological Issues in Metaethics: Moral Particularism vs. Moral Generalism

Metaethics is also concerned with moral epistemology and the nature and process of making moral judgments. A particular issue in this context concerns the scope of moral judgments. Here, two general positions can be distinguished (Dancy, 2017):

- *Moral particularism* holds that each situation is unique and should be evaluated on its own terms;
- *Moral generalism* holds that moral judgements must depend on general moral principles.

The defining difference between a moral particularist and a moral generalist position concerns the use of moral principles. Moral particularism suggests that moral principles may not always be right and should therefore not be seen as the ultimate arbiter of moral judgements. In effect, each situation should be carefully evaluated on its own terms with due recognition of the present particularities. Moral principles may be informative to consider but should not predetermine the moral judgment. In contrast, moral generalism suggests that moral judgments must to some degree depend on moral principles to ensure the rational application of morality. For instance, it could be argued that without general moral principles inconsistent moral judgements would arise over time.

2.2 Normative Ethics

Normative ethics is concerned with setting standards and norms for conduct on a general level. Thus, it engages with general theories of moral behavior such as *consequentialism, deontology*, or *virtue ethics.* The development of normative ethics has been a core focus of Western moral philosophers since ancient times and continues to attract attention. A general distinction that can be made is between *deontic* theories and *virtue* (also known as *aretaic*) theories (Alexander & Moore, 2021). Whereas deontic theories are concerned with guidance and evaluation in relation to choices (e.g., acts, intentions), virtue theories focus on what kind of persons we are and should be. As it stands, there seems to be no dominant theory of normative ethics with surveys of philosophers indicating that belief in consequentialism, deontology, virtue ethics, or other theories is almost evenly distributed (Bourget & Chalmers, 2014).

Consequentialism

Consequentialism denotes the broad position that normative properties depend only on consequences. It is generally informed by the intuition that what counts in this world are the effects that things or actions will have from the present to the future as we cannot change the past. In normative ethics in particular, consequentialism holds that the morality of a choice (e.g., an act or an intention) depends only on the consequences of that choice. Thus, consequentialists must make explicit what makes certain consequences better than others—they must state what is intrinsically of (dis-)value or, put differently, characterize "the Good" and "the Bad." Generally, those choices that maximize the Good and minimize the Bad are then the morally right ones to make (Sinnott-Armstrong, 2019).

Importantly, when taken as defined above, consequentialism is not a very precise position but more like a broad class of more specific moral theories (Sinnott-Armstrong, 2019). The most prominent example is *classic utilitarianism* which holds that an act is morally right if and only if that act maximizes *hedonic* value (i.e., the Good is understood in terms of *pleasure* and the Bad in terms of *pain)* so that the total amount of pleasure for all minus the total amount of pain for all is greatest. Another example is *preference utilitarianism* which holds that the morality of a choice is to be assessed in terms of the desire satisfaction or preference fulfillment it achieves for the greatest number. As such, both, classic and preference utilitarianism advocate for a monist conception of the Good (i.e., there is only one intrinsic value), which can be contrasted with pluralist accounts and their position that intrinsic value may not be easily reduced into a single value. For instance, *ideal utilitarianism* considers the values of beauty and truth as irreducible components of a theory of value in addition to pleasure (Sinnott-Armstrong, 2019). Needless to say, many more consequentialist moral theories exist—in fact, it has been suggested that any plausible moral theory can be translated into a consequentialist theory (Sinnott-Armstrong, 2019).

A common critique of consequentialist moral theories is that they tend to be very demanding. For instance, a literal (or extremist) implementation of classic utilitarianism would require a moral actor to calculate all impacts of an action so as to maximize the expected hedonic value of actions until the end of time. This task seems insurmountable for any human being given our general *cluelessness* about many of the long-term consequences of our actions (Greaves, 2016). Two options for dealing with this recognition are common for consequentialist thinkers. On the one hand, demanding consequentialist moral theories like classic utilitarianism are sometimes treated as an idealization of what moral actors should aspire to with the full recognition that real world decision making is likely to only ever approximate this ideal. On the other hand, less demanding consequentialist moral theories are being developed in an attempt to align more closely with existing "common sense" intuitions about morality.

Deontology

Deontology (from Greek: duty "*deon*" and study "*logos*") denotes a family of normative moral theories which hold that the morality of choices should be judged by appealing to moral rules rather than their consequences. For example, if there is a moral rule against torture, it is always wrong to torture, even if this could help save the lives of millions as a consequence. To distinguish between universal and culture-bound rules, the term *hypernorm* has been introduced to denote rules, norms or principles that are common to different legal, religious, economic and philosophical cultures (Donaldson & Dunfee, 1994). Altogether, deontology stands in contrast to consequentialist moral theories. Consider the following. In consequentialism, "the Good" has priority over "the Right", whereas deontology suggests the opposite, the Right has priority over the Good. This starting point makes deontology generally less demanding than consequentialism as choices must not always be subservient to the Good to count as morally right, rather what counts is to "simply" not break the rules (Alexander & Moore, 2021).

According to Alexander & Moore (2021), two major classes of deontological theories can be distinguished: *agent-centered* and *patient-centered* theories. Agent-centered theories take the perspective of moral agents and suggest agent-relative *obligations* or *permissions* as reasons for action. For example, parents are commonly argued to have special obligations to their own children that are not shared by other people but also the permission to save their own child even at the cost of not saving one or even more other children to whom they do not have any relation. Core to agent-centered theories is the understanding that *agency* in the form of intention (or other mental states) and/or actions is morally relevant and, thus, morality is a deeply personal matter. People's efforts should be directed at "keeping their own moral house in order" (Alexander & Moore, 2021) rather than focus on the overall good produced. Patient-centered theories are *rights-based* rather than duty-based and hold, for instance, that humans have the right to not being used only as a means without consent even if this would produce good consequences overall. Thus, any choice or action that violates this right would be seen as morally wrong. Patient-centered theories often justify rights in agent-neutral terms to make them plausible and gain support.

A common critique of deontology is the so-called paradox of deontology: If violating rules is bad is it not better to have fewer violations? If so, would it not be right to break one rule if this would avert a sufficiently large number of other rule breakings? How to satisfactorily resolve this question remains open to debate (Alexander & Moore, 2021). Suggested answers include the adoption of consequentialism or the speculative development of a novel deontological rationality that explains how to account for the significance of numbers without deference to consequentialist reasoning.

Virtue Ethics

Virtue ethics denotes a third broad perspective on normative ethics that can be distinguished from consequentialism and deontology. Instead of emphasizing the moral imperative of the consequences or duty-bound nature of choices, virtue ethics holds that *virtue* and *vice* are central and fundamental concepts at the heart of morality (Hursthouse & Pettigrove, 2018). In particular, virtues and vices cannot be reduced simply to traits that produce good consequences (as in consequentialism) or as traits being possessed by those who fulfil their duties (as in deontology). Moreover, other normative concepts such as "the Good" or "the Bad" must be explained in terms of virtues or vices. Virtue ethics has a long tradition dating back to Plato and Aristotle in the West and Mencius and Confucius in the East. Today, four distinct forms of virtue ethics are distinguished: a) eudaimonist virtue ethics, b) agent-based and exemplarist virtue ethics, c) target-centered virtue ethics, and d) Platonistic virtue ethics. While it is beyond the scope of this paper to detail the nuances of these different approaches, we can highlight some of the commonalities that characterize virtue ethics as a distinct school of thought in normative ethics. In particular, almost all approaches to virtue ethics recognize *arête* (i.e., virtue), *phronesis* (i.e., practical wisdom) and *eudaimonia* (i.e., happiness or flourishing) as important concepts (Hursthouse & Pettigrove, 2018).

According to Hursthouse & Pettigrove (2018), virtue is a trait of character that is *arête* or excellent. Unlike a mere habit (e.g., being an early riser), it recognizes a well-entrenched disposition of a person. Possessing a virtue means to *be* a certain type of person and to *embody* a certain complex mindset. Thus, being virtuous is always a matter of degree. For instance, virtue ethicists often distinguish between full virtue and continence (in this context to be understood as "strength of will"). Being fully virtuous would entail acting without a struggle whereas a continent person still needs to exert effort to overcome temptations and unvirtuous desires. *Phronesis* or practical wisdom is the capability to make wise choices. It is an important component for the development of virtue as virtues may be actualized in a variety of more or less excellent ways. It is sometimes characterized as that which enables a person to do the right thing in any situation and understand what is truly worthwhile, truly important, and thereby truly advantageous in life. *Eudaimonia*, often translated as happiness or flourishing, refers to a state of "good spirit" and is often argued to be the highest good that humans can achieve. It is associated with states of happiness, wellbeing and human flourishing. Eudaimonia is understood to be the product or even the goal of living a virtuous life.

A common critique of virtue ethics is that it does not provide actionable guidance as it does not propose an account of "right action" grounded in clear moral principles. A virtue ethicist might respond to such a critique by emphasizing that consideration of virtues and vices and axiological notions of good and bad already provide ample guidance for how to live and what kind of person one should strive to become, and because of this, an account in deontic terms such as right and wrong or duty and obligation may not be necessary at all (Hursthouse & Pettigrove, 2018).

2.3 Applied Ethics

Applied ethics is concerned with the practical application of moral considerations to real world situations because general moral theories and methods are often not specific enough to be applied directly to concrete moral problems in a particular domain (Singer, 2019). Application of ethics may therefore lead to new insights which might in turn contribute back to the discourse on ethics in general. Several fields of applied ethics can be distinguished today, for instance, *bioethics* is concerned with ethics in relation to the life sciences, *environmental ethics* is concerned with the ethical dimension of ecological issues, *business ethics* is concerned with practical ethical questions in the organizational and business realm. Also, with the rise of research on advanced machine learning algorithms in recent years, the topic of *artificial intelligence (AI) ethics* has seen a dramatic growth of interest (Kazim & Koshiyama, 2021; Müller, 2021).

One challenge that all fields of applied ethics have to deal with is the variety of competing ethical theories and viewpoints which characterize contemporary ethics. It can hardly be denied that all major ethical theories may provide useful insights that can inform the application of moral considerations. Taking the domain of IS research as an example, Berente et al. (2011) outline how the rigor vs. relevance debate in IS research (e.g., Agarwal & Lucas Jr, 2005) can be understood as an attempt to reconcile deontological considerations about researcher duties (i.e., requirement for rigor) with consequentialist considerations about the outcomes and impact of work (i.e., requirement for relevance). Both perspectives foreground different aspects that may plausibly be taken to be important for evaluating behavior and are, thus, informative. However, this recognition puts applied ethicists in the bind of having to adjudicate between competing ethical theories which all seem to offer some valuable insights. *Moral uncertainty* (MacAskill et al., 2020) is a term which has recently emerged to describe this situation as a decision problem. Explicitly recognizing the uncertainty of beliefs about moral theories may allow for the integration of arbitrary moral theories into a framework of consequentialist decision making under moral uncertainty (MacAskill et al., 2020). Although certainly not a silver bullet, the concept of moral uncertainty opens up new vistas for how disputes in applied ethics may be approached in a practical and pragmatic manner. For instance, Newberry and Ord (2021) outline how the analogy of a *moral parliament* may be used as a novel framing for decision making under moral uncertainty.[1]

Another challenge in applied ethics concerns the question of who or what is worthy of moral consideration and, thus, belongs into our *moral circle* (e.g., only humans, all intelligent animals, all animals, ecosystems, etc.; see Singer, 2011). Although most relevant to a consequentialist understanding of ethics, far reaching shifts in our cultural attitude towards slavery, the role of women in society, or the moral status of animals demonstrate that all plausible moral theories must provide answers to this question in practice.

[1]The moral parliament analogy assumes that decisions are made based on the outcomes of the negotiations and votes of parliament members, each of which advocating for and representing a certain moral theory. Moral uncertainty is represented by assigning parliament members to moral theories proportional to a person's credence in those theories.

3 Ethics in Design

The explicit consideration of ethics in technology design is a relatively recent phenomenon which started to gain more prominence with the advent of increasingly pervasive digital technologies toward the end of the twentieth century (Franssen et al., 2018). A major question in this realm is whether technology can embody values or, put differently, is *value-laden*. On the one hand, there is the neutrality thesis which holds that technology itself is simply a label for neutral artifacts that may be used for good or bad by its users (e.g., Pitt, 2000). On the other hand, there is the increasingly prominent position that technology design is a purposeful process which constructs technological artifacts to fulfill certain functions that, in turn, make it more or less easy to achieve certain goals over others. Thus, design establishes a propensity in technological artifacts to guide and steer behavior, which seems to make it hard to maintain that technology artifacts remain value-neutral (Van de Poel & Kroes, 2014). Given this recognition there has been a trend to view technologies not as a deterministic and self-contained phenomenon but as path-dependent and emergent from a process of choices during design and use (Franssen et al., 2018). This has led to an increase in attention on ethical questions in the design of technical artifacts such as "how to design artifacts in an ethics-conscious manner?", "how to deal with risks and potential unintended consequences of technologies?", or "how to approach the topic of responsibility in technology design?". We briefly discuss each in turn.

3.1 Ethics-Conscious Design Approaches

A range of ethics-conscious design approaches have been developed in an effort to help designers and engineers integrate ethical considerations into artifact design. One of the most prominent approaches stems from the field of computer ethics and is called Value Sensitive Design (VSD). VSD is positioned as "a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process" (Friedman et al., 2008, p. 70). As Friedman et al. (2008) explain, VSD advocates interlinking conceptual, empirical, and technical investigations in an iterative manner. Conceptual investigations concern fundamental questions about the values and stakeholders implicated in a design project. Empirical investigations of the design context with specific consideration for stakeholder perspectives complement and inform conceptual investigations. Technological investigations concern the properties of the employed technologies and either aim to a) elucidate how they may affect values of concern or b) design value-sensitive solutions.

Other approaches for ethics-conscious design include Enid Mumford's ETHICS approach (Mumford, 1995), value-conscious design (Manders-Huits, 2011b), values-inspired design (Purao & Wu, 2013), ethics by design (Brey et al., 2019), participatory design (Bjerknes & Bratteteig, 1995), the use of focus groups for emancipatory purposes (Stahl et al., 2011), or "Design for X" approaches from engineering (Holt & Barnes, 2010). Although none of these approaches should be seen as a panacea for addressing ethical considerations in design, they provide informative starting points and perspectives from which to assess as well as guide specific design projects. For instance, participatory design may be especially applicable in situations when there is already an agreement between involved stakeholders that democratic values are important and less applicable if some stakeholders have a substantially different perspective on the topic (Friedman & Kahn Jr, 2007). Nevertheless, any design project may learn from considering the assumptions that participatory design makes as well as insights that it affords. Thus, becoming familiar with the available ethics-conscious methods and techniques (see e.g., Friedman et al., 2017; Friedman &

Kahn Jr, 2007) and considering their potential benefits and limitations is an important step that any professional interested in *good* design should consider.

3.2 Technological Risks and Unintended Consequences

A general issue with the design of technological artifacts is how to deal with associated risks as well as potential unintended consequences (Franssen et al., 2018). Whereas a risk is generally understood to be the product of the probability of an undesirable event and the effect of that event, an unintended consequence is a potentially large-scale impact that has not been foreseen. Technological risks and unintended consequences are inherent to technology design in a world that may usefully be characterized as a highly complex and rapidly evolving ecosystem of interacting complex adaptive systems (Perrow, 2011). Thus, the history of technological development is ripe with examples of both. For instance, airplanes have risks associated with the long-term stability of their frame and global warming is an unintended consequence of the development and global adoption of fossil fuel powered engines since the industrial revolution. Disconcertingly, recent research has started highlighting that technological risks and unintended consequences might actually pose a serious threat to the future of humanity if human-level AI and machine learning algorithms are developed that do not remain aligned with human values (Christian, 2020; Russell, 2019). Moreover, even without consideration of such forward looking scenarios, the multitude of nuclear weapons on earth (arguably) already pose a significant threat for the future of humanity and illustrate the very real potential moral significance of technological risks and unintended consequences (Ord, 2020).

How to approach technological risks and potential unintended consequences is the focus of ongoing research. For instance, Brey's (2012) *anticipatory technology ethics* is a framework informed by future studies which aims to provide a flexible approach for the ethical analysis of new and emerging technologies. In a similar spirit, Stahl et al. (2010) present the results of the research project ETICA focused on identifying ethical issues of emerging ICT applications and Wright (2011) proposes a framework for an ethical impact assessment of information technology. Beyond individual approaches, an interdisciplinary field concerned with responsible research and innovation (RRI) has been emerging, amongst others, to help address the dangers of technological risks and unintended consequences associated with the design of new technology artifacts (e.g., Jirotka et al., 2017; Owen et al., 2013; Stahl, 2012b). A hallmark of RRI is an emphasis on anticipation and foresight as part of design processes.

Despite such promising developments, dealing with technological risks and unintended consequences remains a hard problem in practice, especially considering the rapid and, as of yet, mostly unregulated speed at which novel digital technologies are being developed and employed in practice (O'neil, 2016; Zuboff, 2015). Compared to the development of physical technologies, digital technologies are generally quicker and less resource intensive to develop, which makes ethical technology assessments proportionally more costly to developers. Yet, digital technologies can generally scale much more efficiently than physical technologies and impact society to an unprecedented degree. In response, many argue for new approaches to socio-technical research focused on artifact assessment (Rahwan et al., 2019) as well as new forms of digital technology governance (Dafoe, 2018; Gasser & Almeida, 2017). For instance, van de Poel (2016) suggests that the introduction of emerging technologies could be treated as social experiments that are carefully monitored for unintended consequences and limit risk exposure through gradual and incremental roll outs. Other researchers highlight the importance of creating incentive structures that are conducive to cooperative behavior between societies and technology designers (Askell et al., 2019). More work along these lines is desirable to limit our exposure to technological risks and unintended consequences of emerging technologies (Bostrom, 2013).

3.3 Responsibility in Technology Design

As already alluded to in the preceding section, the question of responsibility has become an increasingly important topic in technology design. This includes the elaboration of responsibility on multiple levels: responsibility on the individual level, responsibility on the organizational level, responsibility on the policy level, and responsibility on the cultural level—with the acknowledgement that these levels interact and influence one another.

Responsibility on the individual level is mainly concerned with the personal responsibility of designers, engineers, and other professionals in the context of the development of technological artifacts. Here, codes of ethics or codes of conduct are generally used to articulate and establish widely shared and agreed upon responsibilities that professionals in a specific field hold. For instance, the recently updated *Association of Computing Machinery (ACM) Code of Ethics and Professional Conduct* (2018) is an attempt to establish such a code for the computing profession. It articulates general ethical principles, professional responsibilities, professional leadership principles, and code compliance principles aimed at all computing professionals including "practitioners, instructors, students, influencers, and anyone who uses computing technology in an impactful way" (p. 1). In addition to the code itself additional resources such as case studies and procedural guidelines are provided to demonstrate and help with the application of the code (ACM's Committee on Professional Ethics, 2016, 2021). Although codes of ethics seem to be a necessary step for clarifying and communicating the responsibilities individuals have, empirical research has suggested that taken by themselves they may not be effective in affecting behaviors (McNamara et al., 2018). Thus, further work aimed at promoting responsible and professional behavior seems desirable. For instance, interdisciplinary work involving ethicists, educational researchers, psychologists and behavior change experts focused on developing effective and scalable interventions to improve personal capabilities for enacting responsible and professional behavior seems like a promising avenue for future research.

Responsibility on the organizational level is tightly intertwined with responsibility on the personal level and often associated with the Problem of the Many Hands (PMH) which refers to the problems associated with individual accountability and the ascription of responsibility in collective settings (Franssen et al., 2018). Here, the question becomes how can organizations be designed and managed in a way that responsibility is taken up as a priority and not avoided or unintentionally diffused and lost between the cracks? Whole research streams such as corporate social responsibility have been devoted to this question in general (e.g., Lindgreen & Swaen, 2010) but more recent developments include a strong focus on articulating specific guidelines for responsible technology development in the context of emerging AI ventures—so called AI Ethics (Hagendorff, 2020; Kazim & Koshiyama, 2021). However, researchers have highlighted the challenge associated with ensuring responsible behavior in the business sector and emphasized the need for more comprehensive solutions such as thoughtful and informed regulation (e.g., Askell et al., 2019; Mittelstadt, 2019).

Responsibility on the policy level is concerned with the question of how laws and regulations can be used to help foster responsible behavior across scales. Currently, all major governments of the world are deeply considering how to regulate emerging technologies and in particular novel AI ventures (e.g., European Commission, 2020, 2021; The National Artificial Intelligence Initiative, 2021) and academia is increasingly aiming to inform these efforts through emerging research fields such as AI Governance (e.g., Dafoe, 2018) and RRI (e.g., Stahl, 2011; Stilgoe et al., 2013). Given the unprecedented challenges associated with emerging technologies (Bostrom, 2013), going forward it should be expected that an increasing amount of resources and attention will be directed to policy relevant efforts.

Finally, responsibility on the cultural level is concerned with the narratives, paradigms and logics which encompass and frame all preceding levels. Here, an increasingly prominent view is that individual and societal development can be effectively framed in terms of an ongoing cultural evolution of which technology is an important part (Richerson & Christiansen, 2013). Intriguingly, cultural evolution scholars have recently started to consider cultural evolution itself in terms of an applied and interventionist field aimed at the conscious design of sustainable and prosocial cultural systems (Wilson et al., 2014). Building on such developments, future research may look to cultural evolution as a unifying framework for responsible technology development across levels (e.g., Rahwan et al., 2019; Ronfeldt & Arquilla, 2020; Wilson & Gowdy, 2013; Wilson et al., 2014).

4 Ethics in Science

In science, ethics is a topic that has garnered significant attention after large-scale scandals of unethical behavior by scientists became known during the twentieth century (Fisher & Anushko, 2008). In particular, prisoner experiments by the Nazi scientists and the now infamous Tuskegee Syphilis study by American scientists made it clear that scientists could not always be trusted to behave in ways that the general public would deem ethical. In response, national and international guidelines for biomedical research were established around three general ethical principles (Fisher & Anushko, 2008, p. 96):

- *Beneficence*: the obligation to maximize research benefits and minimize research harms;
- *Respect*: the responsibility to ensure that research participation is informed, rationale, and voluntary;
- *Justice*: the obligation to ensure the fair distribution of research benefits and burdens across populations.

Over time, similar guidelines and principles have been established for the social sciences. Today, most developed and increasingly also developing nations have established mechanisms such as Institutional Review Boards (IRB's) at universities to oversee and ensure adherence to these guidelines and principles. However, as already alluded to in the section on ethics in design, recent developments toward more powerful emerging technologies are starting to stretch the limits of guidance and expertise that IRB's can reasonably provide. More flexible approaches such as discourse ethics (Mingers & Walsham, 2010) with a focus on proactive anticipation and discussion of ethical issues have been suggested as a possible way forward (Stahl et al., 2019).

Another contentious question in science is the inherently ethical concern about the role and responsibilities of scientific fields in relation to societies. The value-free ideal of science holds that the sciences have the right and, for some, even the duty to engage in creative inquiry without being overly considerate of the ethical consequences of such work. Given this perspective, science is best left free of value considerations as to remain a bastion of objectivity and authority on truth about the universe (e.g., Hudson, 2016). In contrast, opponents of this view have forcefully argued that science such as any other human activity is saturated with values (e.g., any acceptance or rejection of a hypothesis requires value judgments about the available evidence) and, ultimately, always a reflection of the values of the involved actors (e.g., Churchman, 1995; Douglas, 2009; Ulrich & Reynolds, 2010). To some degree the disagreement between these two camps may be traced back to differences in metaphysical and epistemological assumptions about the universe but are also at heart, maybe ironically, differences in value judgements about the proper role of the sciences in society. Are science fields best kept solely about the reliable production of knowledge concerning the behavior of the universe or better explicitly directed to aid in the responsible development of humanity's future? Different answers in the form of competing scientific paradigms are likely to emerge as the future unfolds.

5 Ethics in IS Practice

Given the history and development of the IS field at the intersection of management, organizational and computer science, the consideration of ethics in IS practice falls under the purview of business ethics as well as disciplinary codes of ethics and professional conduct such as the one developed by the ACM Code 2018 Task Force (2018), which was already introduced in the section *Ethics in Design*. Business ethics is generally conceived of as "the study of the ethical dimensions of the exchange of goods and services, and of the entities that offer goods and services for exchange" (Moriarty, 2021). Similar to other applied ethics fields, business ethicists have adapted general moral theories such as consequentialism, deontology, and virtue ethics to the business domain and considered and developed their implications. To elaborate these challenges, we draw on the substantial expertise and experiences from the authors to flesh out two key challenges for ethics in IS practice: *a lack of skills and resources* and *value conflicts*.

5.1 Challenge: Lack of Skills and Resources

Considering the speed at which new digital technologies and computing applications are being rolled out, staying up to date on the latest developments in the computing field is a common challenge almost any IS practitioner can relate to. To illustrate the dilemma, we refer to the famous quote by Lewis Carroll (1899): "Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!" Given such stringent demands for adaptation and lifelong learning, it is not surprising that not all IS professionals are always able to make decisions that are not only ethically directed but also competent, based on the most suitable technologies, and backed by the best available evidence. However, given the technological risks and potential unintended consequences of many IS projects today, this situation is increasingly untenable.

As an example, consider the case of the use of IBM's Watson to improve cancer care (Strickland, 2019). The aim here was to do something good, that is, improve health care and benefit patients. The system, however, despite having been deployed worldwide, has been criticized for relatively poor performance (Ross & Swetlitz, 2017). One criticism is that the data set used to train the system was biased to the specific conditions that were present at the hospital where the system was initially developed. Thus, without implying any kind of malicious intent, it appears that the developers of the system either lacked the skills or resources to avoid rolling out and selling a potentially dangerous solution. This is despite a clear warning in the ACM Code of Ethics and Professional Conduct that "Extraordinary care should be taken to identify and mitigate potential risks in machine learning systems. A system for which future risks cannot be reliably predicted requires frequent reassessment of risk as the system evolves in use, or it should not be deployed" (ACM Code 2018 Task Force, 2018, p. 6).

Building on the preceding example, we must also recognize that businesses and other organizations are often not designed to enable practitioners to make *right* or *good* but *profitable* decisions. Genuine and deep concern for corporate social responsibility, that actually empowers employees to build up the skills and resources necessary to make difficult ethical decisions, is (despite increasing attention and efforts, at least from an anecdotal perspective) still very hard to come by. Similar to the predicament of the IS professional, organizations are often pressured by relentlessly ongoing change and economic narratives focused on competition and growth as maxims for business.

Against this backdrop, a lack of skills and resources for ethical concerns is likely to be a formidable challenge for ethics in IS practice at least in the near future. It seems fruitful and important to investigate ways in which ethical and competent behavior can be supported by identifying and developing effective tools, teaching formats, work paradigms, and other supporting infrastructure, always informed by already existing mechanisms and interventions such as codes of conduct, ethics courses at universities, or professional training.

5.2 Challenge: Value Conflicts

Another important challenge for ethics in IS practice is tied to the general observation that in modern societies a broad diversity of value systems exist (e.g., Awad et al., 2018). A logical consequence of this observation is that the presence of multiple value systems can precipitate potential tensions around *conflicting values* and how to deal with them. For instance, the moral machine experiment by Awad et al. (2018) highlights that people with different cultural backgrounds have different moral intuitions about who to save or not to save in a thought experiment about an unavoidable accident involving an autonomous car. Thus, major ethical questions regarding the development of algorithms for autonomous cars arise. Should moral intuitions be taken as authoritative and algorithms be tailored to the cultural norms of a country? Is an autonomous car manufacturer allowed to simply choose what kind of algorithm it wants to implement? Or is there one true and right way to design the algorithm that all autonomous car manufacturers ought to converge on? And how many resources do we need to invest to test and ensure that developed algorithms actually align with the goals that the developers intended?

Finding practical ways of dealing with such ethically charged questions around value conflicts is likely to be a major concern for many IS practitioners going forward. The more we digitalize the world and try to manage it with algorithms, the more we need to become explicit about our own value systems and how these relate to others. Extant research on ethics can inform such efforts but ultimately the IS professionals must find a way to consider these lessons to make informed and responsible decisions before designing and implementing solutions. For example, in the section *Applied Ethics* we have already presented some recent ideas around the notion of moral uncertainty, which could provide a useful reference point for any work that is faced with value conflicts and wants to approach them with explicit consideration of different moral theories. Furthermore, the methods and techniques mentioned in the section *Ethics-Conscious Design Approaches* provide a more design-oriented perspective on how to engage with values in design. In addition, Habermas's discourse ethics (Mingers & Walsham, 2010) should be seen as another informative reference point, as it provides a set of procedural guidelines for dealing with value conflicts through open discourse and argumentation. In a similar vein, Ulrich's (2006) critical pragmatism informed by his work on critical systems heuristics (e.g., Ulrich

& Reynolds, 2010) offers the dialectic application of boundary critique as a means for surfacing and engaging with conflicting viewpoints and value systems. In sum, all of these concepts and approaches may offer useful insights when dealing with value conflicts in IS practice but should not be seen as a panacea. Dealing with value conflicts can become a formidable challenge and may require deep engagement, expertise, and many iterations to get "right".

6 Panelists Perspectives on Ethics in Design Science Research

Integrating the preceding discussions, what can now be said about the specifics of ethics in DSR? In the following, we will first introduce and then present individual viewpoint statements by four distinguished DSR scholars, namely, Matti Rossi, Monica Chiarini Tremblay, Sandeep Purao, and Shirley Gregor, who participated in a panel to discuss the role of ethics in DSR at DESRIST 2020. The panel was organized and coordinated by Alexander Herwix and Amir Haj-Bolouri who join the panelists in offering an individual viewpoint. Together, the set of viewpoints offer a broad and diverse set of perspectives to illuminate how different scholars have experienced the challenges related to ethics in their scholarly journey so far.

6.1 Matti Rossi

Matti Rossi is a professor of IS at Aalto University School of Business. He is a past president of the Association for Information Systems. He was the winner of the 2013 Millennium Distinction Award of Technology Academy of Finland for open source and data research. He is one of the co-authors of the seminal MIS Quarterly paper on Action Design Research (ADR) (Sein et al., 2011).

Personal Viewpoint

Matti Rossi emphasizes the importance of ethics for any kind of design work as well as the study of designed artifacts. This stance is reflected in his affiliation with the Scandinavian Tradition in IS development, which advocates for a democratic and stakeholder value-oriented perspective on IS and organization design. The Scandinavian Tradition is one of the precursors of the participatory design approaches mentioned earlier. The key insight from the Scandinavian Tradition is that 'what makes the systems work' is a good understanding of the actual work processes, which allows to fit the data processing needs of the managers with the information needs of the workers (Bjerknes & Bratteteig, 1995; Ehn & Kyng, 1987). From the design perspective this means that the users should have power to influence the decisions about system functionality (Bjerknes & Bratteteig, 1995) and that there should be explicit design goals for improving the work situation. From an ethics perspective this is clearly a form of value-oriented design, but it is important to note that there is no need for perfect stakeholder agreement on issues, rather the conflicts between stakeholders should be seen as a necessary means to uncover all of the involved values and goals.

He stresses the underlying philosophy of trying to ensure that DSR is performed in a way that is ethically sound and benefits all the parties involved, not just the researchers or managers. He believes that if researchers take such recommendations and principles seriously, it can help to gain the trust of the practitioners, design more ethically sound systems, and generate more appropriate design principles.

In his work, he has been faced with choices that he perceived to have deeply ethical components. For instance, he and his colleagues were approached by an arms development company that wanted to collaborate on a DSR project. Ultimately, they decided against this cooperation pointing to ethical concerns about the potential consequences of the project. So, even before the DSR project was initiated, ethics was perceived to be a relevant factor. In terms of recommendations for research practice, he advocates a democratic and stakeholder-oriented approach to DSR. Looking forward, he urges the IS community to take the increasingly pervasive role of IS in society more seriously and investigate it from ethically informed viewpoints that help us navigate towards positive futures and identify the negative futures to be avoided.

6.2 Monica Chiarini Tremblay

Monica Chiarini Tremblay is a Professor of Business Analytics at William & Mary's Raymond A. Mason School of Business. She is an active DSR scholar with much of her research focused on data analytics and technology adoption and sustainability in the healthcare context. Her work has touched on topics that impact vulnerable communities such as foster children, cancer patients, and veterans. Thus, she has to consider the three general ethical principles of beneficence, respect, and justice (Fisher & Anushko, 2008, p. 96) concerning the vulnerable populations that provide the context for her research.

Personal Viewpoint

Monica Chiarini Tremblay emphasizes that ethical concerns occur throughout the DSR process. Many choices can have severe consequences and not all of them are naturally good. For instance, as social media demonstrates, even good design can have adverse effects. Thus, she suggests that as DSR scholars we have the responsibility to consider the ethical consequences of our actions at every step of the DSR lifecycle. She illustrates this point with an anecdote where she (as a reviewer) felt compelled to reject a paper based on its perceived ethical implications, specifically *beneficence*. The proposed artifact, which predicted microloan default risk based on social media behavior, would negatively impact precisely those who most need microloans. In terms of recommendations for research practice, she highlights several examples in her own work.

First, early in her career, she collaborated with a critical theorist to study the emancipatory ways Information Systems researchers can use focus groups to collect data (Stahl et al., 2011). This work has guided much of her design science work—directing her to consider whether all relevant and identifiable stakeholders are represented and identify opportunities to emancipate both the researcher and the participants.

A second example is a recent study where she and her colleagues were using data from a foster care agency to illustrate the efficacy of a method that organizes data to improve the transparency and performance of machine learning models. One particular piece of data was a perfect example of the effectiveness of their artifact but also highlighted a problematic inefficiency of that foster care agency. The team decided to use a less compelling illustration of the method. Foster care agencies are notoriously under-funded and over-burdened by federal and state regulations. Often data can be weaponized to support competing political views, thus, care should be taken when communicating research results. Though it was unlikely that stakeholders in foster care would read an article in an academic IS journal, it was still important to communicate carefully to reduce the chance of unintended consequences for a cooperating research subject (in this case the agency).

Altogether, she highlights how one must take ethical considerations seriously at every step of the design science cycle. For example, her colleagues used action design science research to design a mHealth application for lung cancer care (Tremblay et al., 2020). The mHealth application utilized service-dominant logic to guide the design of a mHealth app that creates value for clinicians, patients, and caregivers. Stakeholder involvement is paramount in the design phase, but the author team debated the *beneficence* of involving lung cancer patients in the early stages. Lung cancer patients are often very ill. They are not in a mental state conducive to novel conceptualizations, such as imagining the requirements for the design and development of an application. The medical literature advocates taking exceptional care when incorporating patients at a vulnerable time in their lives (i.e., terminally ill patients); it is often preferable to refer to proxies instead (Reid, 2009). The research team worked with clinicians, clinician associations, and patient advocacy groups to collect the requirements and then create the beta version of the application. The team then consulted with healthcare policy and cancer research colleagues to determine if it was appropriate to conduct focus groups with patients to evaluate the beta version of the mHealth app. With the guidance of these experts, the team designed the focus groups that treated patients with *respect* by 1) articulating their situation in an objective and non-judgmental climate, 2) motivating them to consider how being involved in the focus group with other cancer patients could be an interesting escape from normality, and 3) making them aware that the information they provide might help in future planning and delivery of care to others (Reid, 2009). This thoughtful research design paid off and contributed significantly to a successful research project.

6.3 Sandeep Purao

Sandeep Purao is Trustee Professor at the Information and Process Management Group at Bentley University, and Visiting Professor at Agder University, Norway. His current research focuses on the design of technology for the social good, and the sciences of design with research projects that deal with the homeless, the elderly, and domains such as fake news. He is also one of the co-authors of the seminal MIS Quarterly paper on ADR (Sein et al., 2011). He is the recipient of the DESRIST Lifetime Achievement Award in 2014.

Personal Viewpoint

Sandeep Purao argues that ethics and responsible behavior are an integral part of DSR and a prerequisite for effective outcomes. He views ethical behavior through two perspectives, one focused on the actions themselves (i.e., a deontological approach to ethics), the other focused on the consequences of the actions (i.e., a consequentialist approach to ethics). His approach to reconcile the two perspectives involves evaluating the actions and behaviors of human actors by examining the personal circumstances and context of the actor. Such a perspective can provide a basis to understand visible conflicts between the actors, and trace these, not only to the potential value conflicts but also how these values are formed by the actors.

For him, this contextual perspective of ethics requires a self-reflective mindset that is open to criticism and feedback. For instance, he proposes that IS scholars should be more reflective about the research problems they choose, and become more aware of possible criticism of outcomes they report. He further suggests that DSR scholars should become more concerned about the distal and unintended consequences of their work that remain difficult to appreciate because of the limits on the duration of research projects. In an exposition of some of his own research (Purao & Wu, 2013), he highlights the inherently ethical nature of DSR by emphasizing that the novel IT artifacts that DSR scholars generate can enhance or reduce the potential for action (echoing the quote from van den Hoven in the opening passage of this paper: "*[designs and artifacts] are all formidable* *shapers of the world we inhabit and in which we acquire our beliefs, decide and act, expect, feel, and hope"*; Maedche, 2017, p. 298).

Accordingly, his perspective about the consideration of values and ethics in the pursuit of DSR is inspired by, and most closely aligns with ideas related to value-sensitive design (Friedman et al., 2008) that is part of the stream of work related to ethics-conscious design (see Section 3.1 earlier). This perspective encourages the DSR scholars to "frontload" considerations of ethics in their research efforts, instead of relegating these to later generations of scholars who would need to sift through the impacts of their work with the lens of unintended consequences (see Section 3.2 earlier). He acknowledges that such a stance can be particularly difficult for DSR scholars who must assume an anticipatory stance (see the ideas from Brey (2012), explored in Section 3.2 earlier). He, therefore, formulates his recommendations for research practice for DSR scholars by urging them to consider work on ethics in design and engineering to become more conscious about how to integrate "values we value" (ethics) into design efforts (Purao & Wu, 2013).

In addition to considering the impact of their research efforts on potential users, he suggests that the best reason for DSR scholars to care more about ethics should be self-interest. As one example, he points to work in the direction of mitigating the spread of fake news by developing a more nuanced understanding of how it can spread (Purao et al., 2021). He argues that without adequate consideration of ethics in their pursuit of research, DSR scholars run the risk of creating futures they themselves may not want to live in.

6.4 Shirley Gregor

Shirley Gregor is a Professor Emerita at the Australian National University. Her research interests include artificial intelligence, human-computer interaction and the philosophy of science and technology. She is the co-author of numerous seminal papers on the nature of design theory and how to publish DSR for maximum impact (Gregor et al., 2020; Gregor & Hevner, 2013; Gregor & Jones, 2007). In 2017 she was awarded a DESRIST Lifetime Achievement Award. Of particular relevance to the panel is her service 2007-2008 as the Director of the Professional Standards Board of the Australian Computer Society.

Personal Viewpoint

Very broadly speaking, Shirley's personal views on ethics follow the idea that an individual should avoid doing harm to others and with Kant's categorical imperative that one should act in accordance with rules that one is willing to recognize as holding universally for everyone: for example, do not be untruthful. Kant's view is a deontological normative theory as described earlier.

However, Shirley's experiences as a software engineer in the early stages of her career, her time as Director of the Professional Standards Board of the Australian Computer Society and her work on DSR projects have all pointed her to the importance of also considering *applied ethics*. She recognizes principles such as Kant's categorical imperative are very general and likely not able to give specific guidance in some of the very complex situations of information systems development, where outcomes are uncertain and different stakeholders have very different interpretations of what is right and wrong.

Approaches that Shirley has observed as valuable in practice for providing more specific guidance have commonalities with Habermas's discourse ethics (e.g., Habermas, 1996), particularly, in cases where so-called *substantive* hypernorms cannot be identified and sufficiently justified for actual decision-making in organizations. Habermas's discourse ethics introduces *procedural* hypernorms that can be used as a starting point for justifying substantive hypernorms, though moral reasoning processes. The basic idea is that the validity of a moral hypernorm cannot be justified by one individual alone, but should be established in a process of argumentation amongst involved individuals. This approach appears to be of practical value and gives backing for some of the procedures that are observed in information systems development, human-centered design and in codes of good conduct for computer professionals.

Considering applied ethics further, Shirley Gregor emphasizes the importance of ethical behavior through the lens of professional behavior and associated codes of ethics. Professional behavior is deemed particularly important for DSR scholars who often work at the frontier of technology where potential negative outcomes and externalities are always on the cards. She suggests that in such situations "doing things right" (e.g. meaningful involvement of stakeholders, testing software adequately) is as important as "doing the right thing" (i.e. aiming at goals that benefit society).

Shirley has seen cases on DSR projects where researchers did not act professionally and where this behavior could be seen as unethical. In one case a project leader undertook a workshop to design the interface for a new system, although his experience in human-computer interaction was negligible. Experts in the field outside the project could see that the resultant design was unsatisfactory, but the industry partners could not. Other project team members were put in the very difficult position of deciding whether to speak up about the problem or not. The case was particularly difficult because of power imbalances amongst the researchers. Sadly this case is not unusual and IS researchers, along with professionals in many fields, may find themselves having to choose between doing things right and suffering repercussions in their employment.

6.5 Alexander Herwix and Amir Haj-Bolouri

Alexander Herwix is a doctoral candidate at the University of Cologne in Germany and currently mainly concerned with the theoretical and practical advancement of DSR. Amir Haj-Bolouri is an Associate Professor at University West in Sweden and focuses on questions and phenomena concerning design, immersive technologies (e.g., Virtual Reality, Augmented Reality), philosophy, and learning.

Personal Viewpoint

Alexander Herwix and Amir Haj-Bolouri strongly advocate for the centrality of ethical considerations for any type of research (Herwix & Haj-Bolouri, 2021) and DSR in particular (Herwix & Haj-Bolouri, 2020). At heart, all decisions that we are making are either implicitly or explicitly a reflection of our values (Ulrich, 2006). As such, an understanding of ethics is paramount for the assessment of IS research and can help to improve its relevance and impact (Herwix & Haj-Bolouri, 2020, 2021). In particular, they argue that we need to be explicit about the values we consider to be the motivation of our work. But we also need to follow through in terms of measuring our progress towards achieving them when assessing our work. For instance, if we are seriously interested in contributing to sustainability, we need to make explicit what exactly we mean by that and how much sustainability our work helps to create for each dollar that is spent on it. The same goes for more traditional values like organizational value realization or academic knowledge. Making progress on this front seems especially important to them given that recent reviews have suggested that most papers in IS journals right now only make shallow arguments (e.g., focused only on direct and easily measurable effects or no empirical support at all) for the practical import of their work (Moeini et al., 2019; Spindeldreher et al., 2020). As such, they advocate for clearer guidelines around how to report and communicate our work in an ethically conscious manner.

Altogether, they view ethics not as a fixed set of beliefs but as a pragmatic and continually evolving process of collective inquiry aimed at identifying and justifying *wise* behavior (Grossmann et al., 2020; Ulrich, 2006). As Ulrich (2006) outlines, the boundaries we draw to make sense of the world inherently affect the facts and judgements we make and vice versa. Thus, we must become experts at justifying and documenting the boundary judgements we are making, if we are to hope for moral progress. As such, they view ethics as inherently and intimately related to the practice of doing DSR which is itself concerned with the question of how to improve our actions in the world through the design of useful artifacts (Herwix & Zur Heiden, 2022; Hevner et al., 2004). They agree that in a very real sense, "science and ethics ultimately converge" (Ulrich, 2006, p. 67). Their viewpoint emerges from an engagement with several different bodies of knowledge including pragmatism (e.g., Ulrich, 2006), systems thinking (e.g., Churchman, 1995), effective altruism (e.g., MacAskill, 2015), and cognitive science (e.g., Grossmann et al., 2020) as well as wrangling with the practice of doing DSR.

7 Discussion and Recommendations

Evident in the personal viewpoint statements is the agreement among the panelists that ethics remains a critical consideration for IS research in general, and DSR in particular. The very nature of DSR, aimed at designing innovative artifacts and solving real world problems, may, in fact, need to be *defined* in terms of such value considerations (see also Churchman, 1995; Ulrich, 2006). The panelists further agree, and the preceding review of prior work provides the rationale to point out, that this can be challenging for a number of reasons, including the need to be more reflective (Purao & Wu, 2013), more anticipatory (Brey, 2012), more aware (Herwix & Zur Heiden, 2022), and more informed about the many different perspectives and contemporary debates about the nature of ethics and values (e.g., Singer, 2019).

Beyond these larger concerns, there is one immediate issue (also see the preceding discussions on responsibility in design and the challenges for ethics in IS practice) that the panelists acknowledge: the incentive structures within academia that do not acknowledge nor support mechanisms to surface these considerations. For instance, Tremblay et al. (2018) outline how the prevailing quantification of faculty productivity in interplay with entrenched publication norms seems to—at least to some degree—discourage the conduct of DSR. Restructuring such incentive systems to be more conducive to responsible forms of IS research such as ethics-conscious DSR is a difficult challenge. Any transformation in this space is likely to be a very slow and political process that needs to propagate across disconnected universities and journals. As an example, accreditation guidelines are only updated in multi-year intervals and even then, might not be as responsive to change as would be desirable.

There are, however, some promising moves in this direction with the recognition of thought leadership and societal impact within the AACSB standards formulated in 2020. Although these do not specifically acknowledge the role of values and ethics, phrases such as *societal impact* can provide IS scholars new pathways to consider how their work can be guided by concerns of societal value. In spite of these recent moves, work to transform the system remains limited in scope and effect. Real and measurable changes will require hard work and a sustained effort by many scholars in the IS community and beyond. However, given what is at stake, this panel took the position, as outlined in this report, that responsible forms of IS research such as ethics-conscious DSR ought to become the norm in the IS field. In the following, we offer three concrete recommendations that may help us in realizing this effort: *encouraging value reporting*, *engaging in problem prioritization*, and *improving the research infrastructure*.

7.1 Encouraging Value Reporting

We suggest that there is an opportunity to improve the ethical awareness in DSR (and IS research in general) by promoting mechanisms such as "value reporting." By this, we mean the explicit positioning of papers with regard to the underlying value systems that the authors used or considered. One approach to realizing this goal could take the form of a paragraph in the theory development or method section of a paper (or even become a separate section either in the paper or separately as an appendix), where the work is mapped to the main values that the authors want to promote (see Appendix A for two examples). Such value reporting can improve the openness and transparency of research and nudge authors toward explicit reflection about how their work relates to their espoused or professed value system.

Such increased transparency about values would help the IS community to become more self-reflective about the values it promotes, more transparent with regards to potential value conflicts, and also more attuned to novel research opportunities. In particular, it would become much more apparent if non-traditional values such as sustainability or justice have been neglected compared to more traditional values such as business success. In this vein, a recent review indicated that the majority of IS research analyzed used predominantly economic market values (e.g., profit, competition, etc.) to justify their work and neglected other potential values such as sustainability (Spindeldreher et al., 2020). Thus, encouraging authors to reflect and report on the values that provide the basis for their work may help to foster novel research perspectives and more ethics-conscious problem framings.

Moreover, value reporting could create efficiencies by helping to align author intentions and reviewer expectations. Transparent and explicit value reporting by the authors of a paper may help reviewers appreciate their intent in a way that potential ethical concerns can be addressed outright. For instance, if a paper addresses a controversial or ethically delicate topic, having clarity around the motivations and intents of the authors can help editors and reviewers make better informed decisions about the paper and offer suggestions for improvement. We provide examples of value reporting in Appendix A to this paper: one that uses this paper as an example, and the other based on a recent paper from one of the authors.

7.2 Engaging in Problem Prioritization

Considering the impact that the choice of research problems has on the potential relevance of the IS field, we join others in advocating for a more systematic and serious engagement with problem choice, framing, and prioritization (e.g., Becker et al., 2015; Chen, 2011; Gable, 2020; Herwix & Haj-Bolouri, 2020, 2021; Mertens & Barbian, 2015; Purao, 2021; Rai, 2017; Winter & Butler, 2011). Put simply, our work can only be as important as the problem which it is addressing. Here, ethics can provide helpful support to more rigorously assess the importance of research problems against various perspectives on what truly matters (see Herwix & Haj-Bolouri, 2020, 2021). Thus, ethics and applied ethics in particular can provide helpful kernel theory which may inform the discourse around research problem priorities in IS research. In essence, this can provide a more rigorous theoretical grounding for how topic selection decisions are made and contribute to the justifiability and expected impact of the IS field. As such our recommendation is strongly congruent with and supportive of recent calls for a more strategic orientation in IS research (e.g., Gable, 2020; Winter & Butler, 2011).

Building on the above precursors, we encourage future IS researchers to engage in problem prioritization on two levels. First, we encourage individuals and teams of researchers to systematically engage in ethics-conscious problem prioritization in the context of their own work. Several researchers already provide useful starting points for informing such efforts (e.g., Herwix & Haj-Bolouri, 2020, 2021; Herwix & Zur Heiden, 2022; Purao, 2021; Rai, 2017). Second, we hope that the IS community as a whole will proactively support further development of mechanisms and approaches that can help to establish systematic and ethics-conscious problem prioritization as a self-organizing feature of the IS field. For instance, specific sections in our top journals could be dedicated to systematic and ethics-conscious discourse about problem priorities in our field and regular special issues be used to incentivize work on the most pressing topics. Prior work on grand challenges for IS research may provide a useful starting point to inform such efforts (e.g., Becker et al., 2015; Chen, 2011; Mertens & Barbian, 2015; Winter & Butler, 2011).

7.3 Improving Research Infrastructure

Building on and extending the two preceding recommendations, we suggest that there is a general opportunity to systematically work on developing a comprehensive research infrastructure (i.e., the socio-technical ecosystem supporting our research efforts; e.g., Dutton, 2011; Morana et al., 2018) that supports and encourages ethical behavior and work. As Winter and Butler (2011) highlight, the DSR community (as well as the broader IS community) can leverage its expertise to help design and build the research infrastructure (e.g., with the suggestion of possible solutions, the construction and evaluation of prototypes, or the identification of useful design principles) that is necessary to tackle large scale challenges of societal relevance and ethical import. This type of work seems highly promising from an expected impact perspective as research infrastructure developed for the IS field may be exapted to other fields and, thus, contribute to improving the overall research infrastructure at large. For instance, the IS community could develop new citation measures that would help to redefine a research communities notion of academic success to be more appreciative of ethical concerns and shape its development trajectory significantly in the long run (e.g., Fitzgerald et al., 2019).

Moreover, self-reflective work aimed at improving our ability to carry out ethics-conscious and responsible research more effectively and efficiently could create virtuous feedback loops that enhance the quality and legitimacy of our work. For instance, citizen science projects in fields like astronomy demonstrate that large gains in research quality and productivity are possible if socio-technical research infrastructure is leveraged in the right way (Watson & Floridi, 2018). Beyond such advanced technological solutions there is also room for the development of innovative teaching formats that are aimed at increasing ethical awareness and competence (e.g., Berti et al., 2021; Parks-Leduc et al., 2021).

Given such opportunities, we posit that the IS community should use at least some of its resources to self-reflectively look for and develop innovative improvements to the research infrastructure. To support such efforts, assessment criteria and decisions for or against publications could be made more transparent. Today, peer review in IS journals remains an opaque process, where acceptance or rejection decisions and underlying reviews are kept from public scrutiny. To increase the transparency and quality of the peer review process, open peer review practices (Ross-Hellauer, 2017) may help. Further, institutional support for working with preprints can also be considered. For instance, publishing anonymous reviews alongside preprints would improve the communities understanding of current review practices and provide a foundation for systematic attempts to improve review quality and, ultimately, research outcomes. Preprints provide a transparent and, in other fields already widely accepted, mechanism for the open-access publication of research designs and outcomes, which improves the visibility of research findings (Fu & Hughey, 2019) and also enables innovations in the research infrastructure (e.g., Sinapayen, 2021).

8 Conclusion

In this report, we have discussed ethics in relation to IS research in general and DSR in particular from four intersecting perspectives: philosophy, design, science, and IS practice. As we have outlined, each of these perspectives provides important concepts and considerations that can inform our own work at the intersection of these fields (an integrative, fifth perspective). Our work is informed by discussions and ensuing reflections at a panel of senior DSR scholars at DESRIST 2020 and builds on deep insights into the real-world significance of ethics in the context of IS research and practice that each of the panelists offers. We hope that our integrative work and reflections can provide a fruitful starting point for more systematic and deliberate engagement with ethics in IS research and DSR in particular. As we have tried to outline in this paper, our collective experience suggests that our field and all of its stakeholders only stand to benefit from it.

Appendix A: Value Reporting

Value reporting refers to the explicit positioning of papers with regard to the underlying value systems that were used or considered. Thus, the main point of value reporting is to clearly state the main values that the authors want to promote through their work and succinctly justify their work in relation to them. We suggest that this could take the form of a paragraph in the method or discussion section of a paper or even be a separate section in an appendix.

In an effort to provide inspiration for the selection of appropriate values, Table 4 lists an overview of eight stakeholder types with associated core values that IS researchers may wish to support with their work. Importantly, we do not claim that this overview is exhaustive or even comprehensive. It merely aims to provide some inspiration that should be seen as a starting point for a more in-depth investigation of potential values that are useful to consider when planning and conducting IS research.

Table 4: Example Stakeholder Types and Associated Core Values to Consider for Value Reporting

Stakeholder Type	**Core Value**	**Description**
Academic	Knowledge	An academic is interested in advancing knowledge.
Practitioner	Satisfaction	A practitioner is interested in achieving satisfaction with her performance.
Individual	Wellbeing	An individual is interested in personal wellbeing.
Organization	Value Realization	An organization is interested in realizing value for its members.
Business	Success	A business is interested in success within a competitive environment.
Government	Societal Welfare	A government is interested in managing and increasing societal welfare.
Civil Society	Justice	A civil society organization is interested in promoting justice in relation to a particular cause.
Future People	Sustainability	Future people are interested in the long-term sustainability of the human project.

Value Reporting Statement: Example 1 (*this* paper)

This paper has been written mainly with the intention of improving the *value realization* of the IS research community. On a very abstract level we agree with the seminal works on DSR that relevance and rigor are important core values that ought to define IS research (Hevner et al., 2004). In particular, we hold the assumption that if we increase the relevance and rigor of our work that, all else being equal, we improve the value realization of the IS research community. Against this backdrop, we see ethics as an important field of reference that can help to clarify the notion of relevance (i.e., what is important) in a rigorous way (Herwix & Haj-Bolouri, 2020, 2021). Thus, we suggest that promoting a more systematic integration of ethics into IS research and DSR in particular has the potential to advance, both, the relevance as well as the rigor of the IS research community and, therefore improve its value realization.

We aim to achieve a more systematic integration of ethics into IS research by advancing the state of *knowledge* about ethics in relation to IS research. To achieve this goal, we have reviewed the literature on ethics and also engaged a set of senior DSR scholars to reflect about their understanding and professional experience with ethics in the context of their work. We present our results as a discussion of five intersecting perspectives on ethics: philosophy, design, science, IS practice, and the panelists perspectives. We also offer a synthesis in the form of a set of recommendations for the advancement of ethics in IS research. We deem this to be an appropriate approach given an observed lack of engagement with ethics from a holistic and experiential point of view. While our goal is to provide some guidance, we recognize the complexity of the topic of ethics and our limited experience engaging with ethics in its entirety. No contributor to this paper has a professional degree in ethics, yet all of us have substantial experience in engaging with ethical questions related to our professional work. Thus, we claim a contribution to the IS research knowledge base but encourage future work to critique and improve upon our work.

Value Reporting Statement: Example 2 (*a different* paper)

The following value reporting statement corresponds to: Purao, S., Murungi, D., Yates, D. 2021. Deliberative Breakdowns in the Social Representation Process: Evidence from Reader Comments in Partisan News Sites. *ACM Transactions on Social Computing*. Vol. 4, Issue 2, pp. 1-35.

This paper was written with the intention of exploring the *value realization* potential of design science research aimed at the problem of dealing with partisan news sites. The authors assume that open, informed, and well-reasoned dialog is a necessary ingredient for a healthy democracy and that substantial value can be realized if dialog can be shaped to exhibit more of these qualities. It is hypothesized that one consequence of the recent popularity of partisan news sites is a fragmentation of dialog across these sites, which may perpetuate increasingly insulated echo chambers. Thus, the authors argue for a need to better understand the discourse amongst the participants at these sites as a precursor to designing innovative solutions that may open up the echo chambers.

The authors aim to contribute to this goal by examining how readers at these partisan news sites consume comments contributed by other visitors (in addition to the actual news stories). The work examines stories and reader comments at two partisan news sites (Breitbart and Daily Kos) surrounding a recent case (Alabama senator Roy Moore who was accused of sexual misconduct and eventually lost the seat in a special election). The analysis reveals multiple deliberative breakdowns—rhetorical, epistemological and emotional—that appear in the reader comments on both partisan news sites. These results are interpreted as pointing to important problems that DSR scholars can address by designing solutions that can promote more open, informed and well-reasoned dialog towards a healthy democracy.

References

ACM Code 2018 Task Force. (2018). Acm Code of Ethics and Professional Conduct. Retrieved September 22, 2021, from https://dora.dmu.ac.uk/bitstream/handle/2086/16422/acm-code-of-ethics-and-professional-conduct.pdf

ACM's Committee on Professional Ethics. (2016, June 24). *Using the Code*. Retrieved September 22, 2021, from https://ethics.acm.org/code-of-ethics/using-the-code/

ACM's Committee on Professional Ethics. (2021). Proactive CARE for Computing Professionals.

Agarwal, R., & Lucas Jr, H. C. (2005). The information systems identity crisis: Focusing on high-visibility and high-impact research. *MIS Quarterly*, 381–398.

Alexander, L., & Moore, M. (2021). Deontological Ethics. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. Retrieved September 13, 2021, from https://plato.stanford.edu/archives/sum2021/entries/ethics-deontological/

Askell, A., Brundage, M., & Hadfield, G. (2019). The Role of Cooperation in Responsible AI Development. *arXiv:1907.04534 [cs]*. Retrieved October 2, 2019, from http://arxiv.org/abs/1907.04534

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Becker, J., vom Brocke, J., Heddier, M., & Seidel, S. (2015). In Search of Information Systems (Grand) Challenges: A Community of Inquirers Perspective. *Business & Information Systems Engineering*, *57*(6), 377–390. https://doi.org/10.1007/s12599-015-0394-0

Benke, I., Feine, J., Venable, J., & Maedche, A. (2020). On Implementing Ethical Principles in Design Science Research. *AIS Transactions on Human-Computer Interaction*, *12*(4), 206–227. https://doi.org/10.17705/1thci.00136

Berente, N., Gal, U., & Hansen, S. (2011). Ethical implications of social stratification in information systems research. *Information Systems Journal*, *21*(4), 357–382. https://doi.org/10.1111/j.1365-2575.2010.00353.x

Berti, M., Jarvis, W., Nikolova, N., & Pitsis, A. (2021). Embodied Phronetic Pedagogy: Cultivating Ethical and Moral Capabilities in Postgraduate Business Students. *Academy of Management Learning & Education*, *20*(1).

Bjerknes, G., & Bratteteig, T. (1995). User Participation and Democracy: A Discussion of Scandinavian Research on System Development. *Scandinavian Journal of Information Systems*, *7*(1), 73–98.

Bostrom, N. (2013). Existential Risk Prevention as Global Priority. *Global Policy*, *4*(1), 15–31. https://doi.org/10.1111/1758-5899.12002

Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe? *Philosophical Studies*, *170*(3), 465–500. https://doi.org/10.1007/s11098-013-0259-7

Brey, P. (2012). Anticipatory Ethics for Emerging Technologies. *NanoEthics*, *6*(1), 1–13. https://doi.org/10.1007/s11569-012-0141-7

Brey, P., Lundgren, B., Macnish, K., & Ryan, M. (2019). *Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design approach*. SHERPA. Retrieved December 27, 2019, from https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf

Carroll, L. (1899). *Through the Looking-Glass and What Alice Found There*. G.H. McKibbin. https://www.loc.gov/item/00000848/

Chen, H. (2011). Editorial: Design science, grand challenges, and societal impacts. *ACM Transactions on Management Information Systems*, *2*(1), 1–10. https://doi.org/10.1145/1929916.1929917

Christian, B. (2020, October 6). *The Alignment Problem: Machine Learning and Human Values* (1st edition). W. W. Norton & Company.

Churchman, C. W. (1995). Ethics and science. *Systems Research*, *12*(4), 267–271. https://doi.org/10.1002/sres.3850120406

Dafoe, A. (2018). *Ai Governance: A Research Agenda*. Oxford, UK, Future of Humanity Institute. Retrieved August 20, 2019, from https://www.fhi.ox.ac.uk/wp-content/uploads/GovAIAgenda.pdf

Dancy, J. (2017). Moral Particularism. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2017). Metaphysics Research Lab, Stanford University. Retrieved September 7, 2021, from https://plato.stanford.edu/archives/win2017/entries/moral-particularism/

DeLapp, K. (n.d.). Metaethics. *Internet encyclopedia of philosophy*. Retrieved September 7, 2021, from https://iep.utm.edu/metaethi/

Donaldson, T., & Dunfee, T. W. (1994). Toward A Unified Conception Of Business Ethics: Integrative Social Contracts Theory. *Academy of Management Review*, *19*(2), 252–284. https://doi.org/10.5465/amr.1994.9410210749

Douglas, H. E. (2009). *Science, policy, and the value-free ideal* (1st). University of Pittsburgh Press.

Dutton, W. H. (2011). The politics of next generation research: Democratizing research-centred computational networks. *Journal of Information Technology*, *26*(2), 109–119. https://doi.org/10.1057/jit.2011.2

Ehn, P., & Kyng, M. (1987). The collective resource approach to systems design. *Computers and democracy*, 17–57.

Ethics. (n.d.). Retrieved December 8, 2020, from https://dictionary.cambridge.org/us/dictionary/english/ethics

European Commission. (2020). *White Paper on Artificial Intelligence – A European approach to excellence and trust*. European Commission. Retrieved April 24, 2020, from https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission. (2021). *2030 Digital Compass: the European way for the Digital Decade*.

European Group on Ethics in Science and New Technologies. (2018). *Statement on artificial intelligence, robotics and 'autonomous' systems*
ZSCC: NoCitationData[s0] OCLC: 1131097250.

Fisher, C. B., & Anushko, A. E. (2008). Research ethics in social science. *The SAGE handbook of social research methods*, 95–109.

Fitzgerald, B., Dennis, A. R., An, J., Tsutsui, S., & Muchala, R. C. (2019). Information Systems Research: Thinking Outside the Basket and Beyond the Journal. *Communications of the Association for Information Systems*, 110–133. https://doi.org/10.17705/1CAIS.04507

Franssen, M., Lokhorst, G.-J., & van de Poel, I. (2018). Philosophy of Technology. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University. Retrieved September 21, 2021, from https://plato.stanford.edu/archives/fall2018/entriesechnology/

Friedman, B., Hendry, D. G., & Borning, A. (2017). A Survey of Value Sensitive Design Methods. *Foundations and Trends in Human-Computer Interaction*, *11*(2), 63–125. https://doi.org/10.1561/1100000015

Friedman, B., Jr, P. H. K., & Borning, A. (2008). Value Sensitive Design and Information Systems. *The handbook of information and computer ethics.*

Friedman, B., & Kahn Jr, P. H. (2007). Human values, ethics, and design. *The human-computer interaction handbook* (pp. 1267–1292). CRC press.

Fu, D. Y., & Hughey, J. J. (2019). Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *eLife*, *8*, e52646. https://doi.org/10.7554/eLife.52646

Gable, G. G. (2020). Viewpoint: Information systems research strategy. *The Journal of Strategic Information Systems*, *29*(2), 101620. https://doi.org/10.1016/j.jsis.2020.101620

Gasser, U., & Almeida, V. A. (2017). A Layered Model for AI Governance. *IEEE Internet Computing*, *21*(6), 58–62. https://doi.org/10.1109/MIC.2017.4180835

Greaves, H. (2016). Cluelessness. *Proceedings of the Aristotelian Society*, *116*(3), 311–339. https://doi.org/10.1093/arisoc/aow018

Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). The Anatomy of a Design Principle. *Journal of the Association for Information Systems.*

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *32*(2), 337–355.

Gregor, S., & Jones, D. (2007). The Anatomy of a Design Theory. *Journal of the Association for Information Systems*, *8*(5), 312–335.

Grossmann, I., Weststrate, N. M., Ardelt, M., Brienza, J. P., Dong, M., Ferrari, M., Fournier, M. A., Hu, C. S., Nusbaum, H. C., & Vervaeke, J. (2020). The Science of Wisdom in a Polarized World: Knowns and Unknowns. *Psychological Inquiry*, *31*(2), 103–133. https://doi.org/10.1080/1047840X.2020.1750917

Habermas, J. (1996, August 13). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (1st edition). Polity.

Hagendorff, T. (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

Herwix, A., & Haj-Bolouri, A. (2020). Having a Positive Impact with Design Science Research – Learning from Effective Altruism. In S. Hofmann, O. Müller, & M. Rossi (Eds.), *Designing for digital transformation. co-creating services with citizens and industry* (pp. 235–246). Springer International Publishing. https://doi.org/10.1007/978-3-030-64823-7_22

Herwix, A., & Haj-Bolouri, A. (2021). Revisiting the Problem of the Problem – An Ontology and Framework for Problem Assessment in IS Research. *Proceedings of the Twenty-Nineth European Conference on Information Systems (ECIS2021).*

Herwix, A., & Zur Heiden, P. (2022). Context in Design Science Research: Taxonomy and Framework. *Proceedings of the 55th Annual Hawaii International Conference on System Sciences.* https://doi.org/10.24251/HICSS.2022.705

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science In Information Systems Research. *MIS Quarterly*, *28*(1), 75–105.

Holt, R., & Barnes, C. (2010). Towards an integrated approach to "Design for X": an agenda for decision-based DFX research. *Research in Engineering Design*, *21*(2), 123–136. https://doi.org/10.1007/s00163-009-0081-6

Hudson, R. (2016). Why We Should Not Reject the Value-Free Ideal of Science. *Perspectives on Science*, *24*(2), 167–191. https://doi.org/10.1162/POSC_a_00199

Hursthouse, R., & Pettigrove, G. (2018). Virtue Ethics. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2018). Metaphysics Research Lab, Stanford University. Retrieved June 23, 2020, from https://plato.stanford.edu/archives/win2018/entries/ethics-virtue/

Jirotka, M., Grimpe, B., Stahl, B. C., Eden, G., & Hartswood, M. (2017). Responsible research and innovation in the digital age. *Communications of the ACM*, *60*(5), 62–68. https://doi.org/10.1145/3064940

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, *2*(9). https://doi.org/10.1016/j.patter.2021.100314

Lindgreen, A., & Swaen, V. (2010). Corporate Social Responsibility. *International Journal of Management Reviews*, *12*(1), 1–7. https://doi.org/10.1111/j.1468-2370.2009.00277.x

MacAskill, W. (2015). *Doing good better: Effective altruism and a radical new way to make a difference*. Guardian Faber Publishing.

MacAskill, W., Bykvist, K., & Ord, T. (2020). *Moral Uncertainty*. Oxford University Press. https://www.moraluncertainty.com/

Maedche, A. (2017). Interview with Prof. Jeroen van den Hoven on "Why do Ethics and Values Matter in Business and Information Systems Engineering?". *Business & Information Systems Engineering*, *59*(4), 297–300. https://doi.org/10.1007/s12599-017-0476-2

Manders-Huits, N. (2011a). What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics*, *17*(2), 271–287. https://doi.org/10.1007/s11948-010-9198-2

Manders-Huits, N. (2011b). What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics*, *17*(2), 271–287. https://doi.org/10.1007/s11948-010-9198-2

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development? *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 729–733. https://doi.org/10.1145/3236024.3264833

Mertens, P., & Barbian, D. (2015). Researching "Grand Challenges": A "Grand Challenge". *Business & Information Systems Engineering*, *57*(6), 391–403. https://doi.org/10.1007/s12599-015-0405-1

Mingers, J., & Walsham, G. (2010). Toward ethical information systems: The contribution of discourse ethics. *MIS Quarterly*, *34*(4), 833–854.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Moeini, M., Rahrovani, Y., & Chan, Y. E. (2019). A review of the practical relevance of IS strategy scholarly research. *The Journal of Strategic Information Systems*, *28*(2), 196–217. https://doi.org/10.1016/j.jsis.2018.12.003

Morality. (n.d.). Retrieved December 8, 2020, from https://dictionary.cambridge.org/us/dictionary/english/morality

Morana, S., vom Brocke, J., Maedche, A., Seidel, S., Adam, M. T. P., Bub, U., Fettke, P., Gau, M., Herwix, A., Mullarkey, M. T., Nguyen, H. D., Sjöström, J., Toreini, P., Wessel, L., & Winter, R. (2018). Tool Support for Design Science Research—Towards a Software Ecosystem: A Report from a DESRIST 2017 Workshop. *Communications of the Association for Information Systems*, *43*(1), 237–256. https://doi.org/10.17705/1CAIS.04317

Moriarty, J. (2021). Business Ethics. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Fall 2021). Metaphysics Research Lab, Stanford University. Retrieved September 23, 2021, from https://plato.stanford.edu/archives/fall2021/entries/ethics-business/

Müller, V. C. (2021). Ethics of Artificial Intelligence and Robotics. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. Retrieved September 20, 2021, from https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/

Mumford, E. (1995). *Effective systems design and requirements analysis: The ETHICS approach*. Macmillan Education UK.

Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science research in information systems. *Information & Management*, *51*(6), 801–809. https://doi.org/10.1016/j.im.2014.01.002

Newberry, T., & Ord, T. (2021). *The Parliamentary Approach to Moral Uncertainty* (#2021-2). Future of Humanity Institute.

O'neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Ord, T. (2020). *The precipice: Existential risk and the future of humanity*. Hachette Books.

Owen, R., Bessant, J., & Heintz, M. (Eds.). (2013). *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society* (1st ed.). Wiley.

Paradice, D., Freeman, D., Hao, J., Lee, J., & Hall, D. (2018). A Review of Ethical Issue Considerations in the Information Systems Research Literature. *Foundations and Trends® in Information Systems*, *2*(2), 117–236. https://doi.org/10.1561/2900000012

Parks-Leduc, L., Mulligan, L., & Rutherford, M. A. (2021). Can Ethics Be Taught? Examining the Impact of Distributed Ethical Training and Individual Characteristics on Ethical Decision-Making. *Academy of Management Learning & Education*, *20*(1), 30–49. https://doi.org/10.5465/amle.2018.0157

Perrow, C. (2011, October 12). *Normal Accidents: Living with High Risk Technologies*. Princeton University Press.

Pitt, J. C. (2000). *Thinking about Technology: Foundations of the Philosophy of Technology*. Seven Bridges Press.

Porra, J. (2001). A Dialogue with C. West Churchman. *Information Systems Frontiers*, *3*(1), 19–27.

Purao, S. (2013). Truth or dare: The ontology question in design science research. *Journal of Database Management (JDM)*, *24*(3), 51–66.

Purao, S. (2021). Design Science Research Problems . . . Where Do They Come From? In L. Chandra Kruse, S. Seidel, & G. I. Hausvik (Eds.), *The next wave of sociotechnical design* (pp. 99–111). Springer International Publishing. https://doi.org/10.1007/978-3-030-82405-1_12

Purao, S., Murungi, D., & Yates, D. (2021). Deliberative Breakdowns in the Social Representation Process: Evidence from Reader Comments in Partisan News Sites. *ACM Transactions on Social Computing*, *4*(2), 1–35. https://doi.org/10.1145/3450143

Purao, S., & Wu, A. (2013). Towards Values-inspired Design: The Case of Citizen-Centric Services. *ICIS 2013 Proceedings*, 8.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. S., . . . Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y

Rai, A. (2017). Editor's Comments: Avoiding Type III Errors: Formulating IS Research Problems that Matter. *MIS Quarterly*, *41*(2), iii–vii.

Reid, J. (2009). Conducting qualitative research with advanced cancer patients and their families: Ethical considerations. *International Journal of Palliative Nursing*, *15*(1), 30–33. https://doi.org/10.12968/ijpn.2009.15.1.37950

Richerson, P. J., & Christiansen, M. H. (2013, November 1). *Cultural Evolution: Society, Technology, Language, and Religion*. MIT Press.

Roes, F. (2003). Belief in moralizing gods. *Evolution and Human Behavior*, *24*(2), 126–135. https://doi.org/10.1016/S1090-5138(02)00134-4

Ronfeldt, D., & Arquilla, J. (2020). *Whose Story Wins: Rise of the Noosphere, Noopolitik, and Information-Age Statecraft*. RAND Corporation. https://doi.org/10.7249/PEA237-1

Ross, C., & Swetlitz, I. (2017). Ibm pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close [magazine]. *STAT*. Retrieved April 12, 2021, from https://www.statnews.com/2017/09/05/watson-ibm-cancer/

Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research*, *6*. https://doi.org/10.12688/f1000research.11369.2

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Sayre-McCord, G. (2014). Metaethics. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2014). Metaphysics Research Lab, Stanford University. Retrieved September 7, 2021, from https://plato.stanford.edu/archives/sum2014/entries/metaethics/

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, *35*, 37–56.

Sinapayen, L. (2021). The Mimosa Manifesto: a Web Platform for Open Collaboration in Science. Retrieved October 18, 2021, from https://openreview.net/forum?id=S-2p4gucHOu

Singer, P. (2011, April 18). *The Expanding Circle: Ethics, Evolution, and Moral Progress*. Princeton University Press.

Singer, P. (2019). Ethics. *Encyclopedia britannica*. Encyclopædia Britannica, inc. Retrieved October 2, 2019, from https://www.britannica.com/topic/ethics-philosophy

Sinnott-Armstrong, W. (2019). Consequentialism. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2019). Metaphysics Research Lab, Stanford University. Retrieved June 23, 2020, from https://plato.stanford.edu/archives/sum2019/entries/consequentialism/

Spindeldreher, K., Schlagwein, D., & Schoder, D. (2020). How is Information Systems Research Justified? An Analysis of Justifications Given by Authors. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 10.

Stahl, B. C. (2011). It for a better future: How to integrate ethics, politics and innovation. *Journal of Information, Communication and Ethics in Society*, *9*(3), 140–156. https://doi.org/10.1108/14779961111167630

Stahl, B. C. (2012a). Morality, Ethics, and Reflection: A Categorization of Normative IS Research. *Journal of the Association for Information Systems*, *13*(8), 636–656. https://doi.org/10.17705/1jais.00304

Stahl, B. C. (2012b). Responsible research and innovation in information systems. *European Journal of Information Systems*, *21*(3), 207–211. https://doi.org/10.1057/ejis.2012.19

Stahl, B. C., Akintoye, S., Fothergill, B. T., Guerrero, M., Knight, W., & Ulnicane, I. (2019). Beyond Research Ethics: Dialogues in Neuro-ICT Research. *Frontiers in Human Neuroscience*, *13*. https://doi.org/10.3389/fnhum.2019.00105

Stahl, B. C., Heersmink, R., Goujon, P., Flick, C., Van Den Hoven, J., Wakunuma, K., Ikonen, V., & Rader, M. (2010). Identifying the ethics of emerging information and communication technologies: An essay on issues, concepts and method. *International Journal of Technoethics (IJT)*, *1*(4), 20–38.

Stahl, B. C., Tremblay, M. C., & LeRouge, C. M. (2011). Focus groups and critical social IS research: How the choice of method can promote emancipation of respondents and researchers. *European Journal of Information Systems*, *20*(4), 378–394. https://doi.org/10.1057/ejis.2011.21

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Strickland, E. (2019). Ibm Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, *56*(4), 24–31. https://doi.org/10.1109/MSPEC.2019.8678513

The National Artificial Intelligence Initiative. (2021). *The National Artificial Intelligence Initiative (NAII)*. Retrieved September 23, 2021, from https://www.ai.gov/

Tremblay, M. C., Cucciniello, M., Tarricone, R., Porumbescu, G. A., & Desouza, K. C. (2020). Delivering Effective Care Through Mobile Apps: Findings from a Multi-stakeholder Design Science Approach. In S. Hofmann, O. Müller, & M. Rossi (Eds.), *Designing for digital transformation. co-creating services with citizens and industry* (pp. 3–14). Springer International Publishing. https://doi.org/10.1007/978-3-030-64823-7_1

Tremblay, M. C., VanderMeer, D., & Beck, R. (2018). The Effects of the Quantification of Faculty Productivity: Perspectives from the Design Science Research Community. *Communications of the Association for Information Systems*, *43*(1). https://doi.org/10.17705/1CAIS.04334

Ulrich, W. (2006). Critical Pragmatism: A New Approach to Professional and Business Ethics. *Interdisciplinary yearbook for business ethics. v. 1, v. 1*. Peter Lang Pub Inc.

Ulrich, W., & Reynolds, M. (2010). Critical Systems Heuristics. In M. Reynolds & S. Holwell (Eds.), *Systems approaches to managing change: A practical guide* (pp. 243–292). Springer London. https://doi.org/10.1007/978-1-84882-809-4_6

Van de Poel, I., & Kroes, P. (2014). Can technology embody values? *The moral status of technical artefacts* (pp. 103–124). Springer.

van den Hoven, J. (2007). ICT and Value Sensitive Design. In P. Goujon, S. Lavelle, P. Duquenoy, K. Kimppa, & V. Laurent (Eds.), *The Information Society: Innovation, Legitimacy, Ethics and Democracy In honor of Professor Jacques Berleur s.j.* (pp. 67–72). Springer US. https://doi.org/10.1007/978-0-387-72381-5_8

van de Poel, I. (2016). An Ethical Framework for Evaluating Experimental Technology. *Science and Engineering Ethics*, *22*(3), 667–686. https://doi.org/10.1007/s11948-015-9724-3

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Walsham, G. (1996). Ethical theory, codes of ethics and IS practice. *Information Systems Journal*, 13.

Walsham, G. (2012). Are we making a better world with ICTs? Reflections on a future agenda for the IS field. *Journal of Information Technology*, *27*(2), 87–93. https://doi.org/10.1057/jit.2012.4

Watson, D., & Floridi, L. (2018). Crowdsourced science: Sociotechnical epistemology in the e-research paradigm. *Synthese*, *195*(2), 741–764. https://doi.org/10.1007/s11229-016-1238-2

Wilson, D. S., & Gowdy, J. M. (2013). Evolution as a general theoretical framework for economics and public policy. *Journal of Economic Behavior & Organization*, *90*, S3–S10. https://doi.org/10.1016/j.jebo.2012.12.008

Wilson, D. S., Hayes, S. C., Biglan, A., & Embry, D. D. (2014). Evolving the future: Toward a science of intentional change. *Behavioral and Brain Sciences*, *37*(4), 395–416. https://doi.org/10.1017/S0140525X13001593

Winter, S. J., & Butler, B. S. (2011). Creating Bigger Problems: Grand Challenges as Boundary Objects and the Legitimacy of the Information Systems Field. *Journal of Information Technology*, *26*(2), 99–108. https://doi.org/10.1057/jit.2011.6

Wright, D. (2011). A framework for the ethical impact assessment of information technology. *Ethics and Information Technology*, *13*(3), 199–226. https://doi.org/10.1007/s10676-010-9242-6

Zuboff, S. (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

Threading the Needle in the Digital Age: Four Paradigmatic Challenges for Responsible Design Science Research

Outline

Abstract

The rapid development of life-changing information technologies poses a significant societal challenge: finding responsible ways to harness their benefits while mitigating potentially catastrophic consequences. However, our understanding of how to contribute to this imperative is still in its early stages, especially in the face of accelerating technological advances, such as in the field of artificial intelligence. In response, this paper proposes the deliberate development of a responsible design science research (DSR) ecosystem as a promising way forward. This responsible DSR ecosystem aims to leverage the insights and diversity of the information systems (IS) research community, along with perspectives from related fields, to address societal challenges in a responsible and productive manner. The paper presents a conceptualization of responsible DSR, characterized by four paradigmatic challenges: ontological, epistemological, axiological, and methodological. Through a multi-grounded research process involving empirical, theoretical, and internal grounding, key challenges for responsible DSR practice are conceptualized and recommendations for its future development are provided. This paper contributes to IS research by presenting a paradigmatic framework that can provide a common ground for the wide variety of IS research interested in addressing real-world problems, and by offering recommendations that open up new perspectives on the possibilities and limitations of responsible DSR. Overall, this work can be seen as nothing more or less than an invitation to embark on a collective journey to imagine a responsible DSR ecosystem and shape its future for the benefit of humanity.

Keywords: Critical Pragmatism, Design Science Research, DSR, DSR Ecosystem, Responsible DSR, Responsible Innovation.

Bibliographic Information

Herwix, A. (2023). Threading the Needle in the Digital Age: Four Paradigmatic Challenges for Responsible Design Science Research. *SocArXiv*.
https://doi.org/10.31235/osf.io/xd423

Author's contribution

Conceptualization, Methodology, Validation, Investigation, Writing – original draft, Writing – review and editing, Visualization, Project administration.

1 Introduction

Information systems (IS) research that helps our societies to anticipate, reflect on, and responsibly shape information technology (IT) innovations is becoming increasingly important in the digital age[1] (Jirotka et al., 2017). The accelerating pace of IT development and innovation holds many promises but also poseses increasingly significant risks for the wellbeing of our societies and even the planet (Brynjolfsson & McAfee, 2012; Ord, 2020; Vinuesa et al., 2020). The seriousness of such concerns is well illustrated by the fact that a large group of prominent artificial intelligence (AI) researchers and other notable figures recently signed a statement declaring that: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war" (Center for AI Safety, 2023). Put simply, a key societal challenge of our time is not just to innovate, but to ensure *responsible innovation* (RI) (e.g., Owen, Bessant, et al., 2013; Stilgoe et al., 2013; Timmermans & Blok, 2021). In practical terms, this means that we need to establish "a collective commitment of care for the future through responsive stewardship of science and innovation in the present" (Owen, Stilgoe, et al., 2013, p. 36), so that together we can find ways to thread the needle between reaping the benefits and avoiding the potentially catastrophic downsides of emerging technologies for our societies.[2]

However, our understanding of how to implement this societal imperative is still nascent, especially in the face of accelerating technological advances (e.g., Blok, 2023; Consilience Papers, 2022; Harari et al., 2023). The IS research community must continue to raise its game to have any chance of keeping up with and remaining relevant to society as the accelerating pace of development and adoption of increasingly advanced IT changes the game board right under our feet (e.g., Desouza & Dawson, 2023; Dwivedi et al., 2023; Gable, 2020; Ram & Goes, 2021). In particular, a common theme in self-reflective IS research is that we need to become better at channeling the diverse perspectives and insights from the IS community and related fields into strategic (Gable, 2020), inter-, multi-, or transdisciplinary[3] research efforts (Galliers, 2003; Kroeze et al., 2019; Nunamaker et al., 2013; Nunamaker et al., 2017) to more effectively address the grand challenges of our societies in the digital age (Paradice et al., 2019; Ram & Goes, 2021; Winter & Butler, 2011).

In response to this situation, this paper envisions the development of a *responsible design science research (DSR) ecosystem* that is explicitly concerned with facilitating, coordinating, and integrating research so that we can implement and advance RI by addressing the grand challenges of the digital age in a responsible and productive manner.

[1]The term "digital age" is used in this paper to refer to the historical period of increasingly pervasive use of IT that began in the mid-20th century (Castells, 2011).

[2]As such, responsible innovation implies a forward-looking and collective understanding of responsibility that recognizes the increasingly global and potentially compounding and long-lasting effects of our local actions (e.g., Stilgoe et al., 2013). As citizens of industrialized societies, we share some responsibility for a global economic system that causes climate change and violates planetary boundaries. Forward-looking responsibility in this context asks us to "take" responsibility and do what reasonably can be done to rein in this out-of-control system. This understanding contrasts with more traditional, backward-looking views of responsibility, which are primarily concerned with assigning responsibility (and blame) for the local and immediate effects of individual actions (Jonas, 1984; Ulrich, 1994).

[3]In the context of this paper, the term interdisciplinary research refers to research in which a topic at the intersection of two or more disciplines is worked on collectively; the term multidisciplinary research refers to research in which a specific topic (possibly outside the general interest of each individual discipline) is viewed and worked on from different disciplinary perspectives; and the term transdisciplinary research refers to research in which disciplinary boundaries are transcended and topics are worked on based on a collectively defined agenda (Kroeze et al., 2019).

This vision builds on and extends, but also challenges, the mainstream understanding of DSR in IS as a research paradigm that is limited to the construction and evaluation of innovative IT-related artifacts[4] and associated design prescriptions that address real-world problems (e.g., Baskerville et al., 2018; Peffers et al., 2007). Thus, DSR is viewed as complementary to, but different in knowledge interest from (and therefore incommensurable with) behavioral science research (BSR), which seeks to develop and justify theories that explain and predict phenomena (Hevner et al., 2004).

On the one hand, I build on and extend DSR because it explicitly recognizes normativity, transformation, and real-world impact as core dimensions of IS research (e.g., Benke et al., 2020; Herwix et al., 2022; Myers & Venable, 2014), whereas BSR tends to do so less explicitly, often claiming a neutral stance with respect to its own relationship to the world (Douglas, 2009; Hevner et al., 2004). In this sense, DSR is the more appropriate starting point for developing a research ecosystem dedicated to implementing and advancing RI.

On the other hand, however, I challenge the taken-for-granted assumption that it is useful and appropriate to erect paradigmatic boundaries between DSR and BSR because of their different knowledge interests (cf. Hevner et al., 2004; Seidel & Watson, 2020). Drawing on a pragmatist understanding of research as inquiry (Dewey, 1938b; Ulrich, 2006, 2016), I argue that all IS research, regardless of its knowledge interests, follows a similar general, conceptual sequence and pattern of problem solving, in which design and prescription are inherently intertwined with explanation and prediction, and vice versa (cf. Farrell & Hooker, 2012; Herwix & Rosenkranz, 2018).

In a pragmatist view, IS research should be limited only by the needs and requirements of the problem situation at hand, never by predetermined knowledge interests, and thus should be open-minded in the specific methods it employs, and open-ended in the concrete outcomes it seeks to achieve (Dewey, 1938b). To do otherwise would amount to a premature and potentially irresponsible limitation of IS research (Dewey, 1938b). Given this understanding, rigid paradigmatic boundaries that might impede problem solving, such as those imposed by our mainstream understanding of DSR, are generally to be avoided. It also recognizes the normative and transformative nature of all forms of IS research. Since all of our actions are selective, none of them can ever be value-neutral: "Science and ethics ultimately converge" (Ulrich, 2006, p. 67). Overall, a pragmatist perspective emphasizes the need for a unifying perspective on responsible DSR that is not restricted to predefined knowledge interests.

As a first step towards realizing the vision of a responsible DSR ecosystem, I propose a framework for responsible DSR that can serve as a "super-methodology" (Nunamaker et al., 1990) for coordinating the responsible DSR ecosystem. Specifically, the framework provides a common foundation for the responsible DSR ecosystem by characterizing the essence of its practice in terms of four fundamental *paradigmatic challenges*:

- *Ontological challenge:* Making warranted boundary judgments,
- *Epistemological challenge:* Using, producing, and growing a body of useful knowledge,
- *Axiological challenge:* Managing conflicting values and positions, and
- *Methodological challenge:* Developing, selecting, and integrating a portfolio of practices.

[4]The term artifact is used in this paper in the spirit of Simon (1996, pp. 6–7), who viewed artifacts as those systems or things that are specifically adapted to serve human purposes. For example, a piece of wood becomes an artifact when it is turned into a spear. In the context of IS research, several types of artifacts have been distinguished such as IT artifacts, IS artifacts, information artifacts, technology artifacts, and social artifacts (e.g., Lee et al., 2015; Orlikowski & Iacono, 2001).

The framework is multi-grounded (Goldkuhl & Cronholm, 2010, 2018) through a research process that combined a) *empirical grounding*, provided by the analysis of 41 in-depth interviews with a diverse set of DSR stakeholders from the IS research community (27 highly experienced academics, 11 early career academics, 3 PhD students), with b) *theoretical grounding*, ensured by the theoretical matching of emerging conceptions with insights from a wide range of scholarly literature, and c) *internal grounding*, driven by ongoing evaluations of theoretical cohesion. The research question I investigated was: *What are key challenges for responsible DSR in practice and how can we improve the status quo?*

I highlight two ways in which this work contributes to IS research. First, the framework for responsible DSR can serve as a useful boundary object (Star & Griesemer, 1989), bridging different perspectives on IS research and enabling a collective research program concerned with supporting and implementing responsible innovation. It is abstract enough to be applicable to the broad diversity of problem- or solution-oriented IS research (Majchrzak et al., 2016; Rai, 2017a, 2019; Van de Ven, 2018), but also practicable enough to be informative and helpful to individual IS researchers in practice. Second, my discussion of specific recommendations for supporting the development of a responsible DSR ecosystem connects the framework to the broader academic literature and points to many promising avenues for future work.

Overall, this paper opens up new and thought-provoking perspectives on the possibilities and limitations of responsible DSR. In doing so, it contributes to a growing stream of IS research arguing that it is time to break out of the narrow design science box (Nunamaker et al., 2013) in which DSR has historically tended to be confined (e.g., Alter, 2012; W. Kuechler & Vaishnavi, 2008; Levy & Hirschheim, 2012; McKay et al., 2012). More strategic, inter-, multi-, and transdisciplinary research efforts are needed to address the grand challenges and most pressing needs of our societies in the digital age (e.g., Desouza & Dawson, 2023; Gable, 2020; Nunamaker et al., 2017; Paradice et al., 2019; Winter & Butler, 2011). A responsible DSR ecosystem is proposed as part of the answer we have been looking for.

The remainder of the paper is structured as follows. First, I outline the conceptual background of responsible DSR. Second, I detail the specifics of the research approach. Third, I present the framework for responsible DSR practice. Fourth, I suggest four actionable recommendations that can help us develop a responsible DSR ecosystem. Fifth, I discuss the findings and results in terms of contributions and limitations. Sixth, I conclude the paper with a call to action to participate in the development of a responsible DSR ecosystem.

2 Background on Responsible Design Science Research

2.1 From Design Science Research to Responsible Design Science Research

DSR has become a well-established research paradigm within the IS research field that focuses on constructing and evaluating artifacts and prescriptions for their design to address specific real-world problems or opportunities (e.g., Baskerville et al., 2018; Gregor & Hevner, 2013). It is often viewed in contrast to BSR, which seeks to develop and justify theories that explain and predict phenomena, and is thus understood to have a different knowledge interest (Hevner et al., 2004; Seidel & Watson, 2020). As such, DSR and BSR have been positioned as complementary paradigms that can reinforce and support each other, but ultimately remain incommensurable (Hevner et al., 2004; March & Smith, 1995). In this mainstream view, DSR would use BSR theories to design novel IT artifacts, which in turn could be studied by BSR to develop new theories that could even better inform future DSR, and so on. Or, in the words of Hevner et al. (2004, p. 80): "Truth informs design and utility informs truth".

However, despite the intuitive appeal of this mainstream view, the sharp distinction between DSR and BSR has proven problematic in the complex reality of IS research practice, where design and prescription are inherently intertwined with explanation and prediction, and vice versa (cf. Baskerville et al., 2015; Farrell & Hooker, 2012; Herwix & Rosenkranz, 2018). In particular, the mainstream view has limited our understanding of what DSR can be by restricting it to a narrow understanding that focuses primarily on the development and evaluation of IT artifacts and prescriptions for their design, rather than a broader understanding that focuses on all aspects of IS design (cf. Alter, 2012; W. Kuechler & Vaishnavi, 2008; Levy & Hirschheim, 2012; McKay et al., 2012; Nunamaker et al., 2013; Purao et al., 2008). In doing so, it has also obscured the potential of DSR to serve as a "super-methodology" (Nunamaker et al., 1990) for systematic programs of IS research (Nunamaker et al., 2013), severely limiting the ability of the IS research community to respond to the resounding calls to more systematically address the grand societal challenges of our time (e.g., Desouza & Dawson, 2023; Gable, 2020; Winter & Butler, 2011). This is particularly problematic in the face of rapidly advancing IT developments, which will require us to make significant collective efforts to ensure their responsible embedding in society, lest we risk undesirable or even catastrophic outcomes (e.g., Jirotka et al., 2017; Russell, 2019; Stilgoe et al., 2013).

Altogether, this situation leads us to argue that we need to move beyond the mainstream view of DSR and towards the development of a *responsible DSR ecosystem* that can enable the IS community to address the grand challenges of the digital age in a responsible and productive manner. This implies two key developments for DSR: (1) a broadening in scope from a research approach for individual research projects to a "super-methodology" for coordinating an entire research ecosystem, and (2) a deeper integration of considerations of responsibility into our work.

First, broadening the scope of DSR from individual research projects to a "super-methodology" for coordinating an entire research ecosystem is a promising step to meet the growing calls for IS research to more systematically address the grand challenges facing our societies (e.g., Boh et al., 2023; Davison et al., 2023; Desouza & Dawson, 2023; Gable, 2020; Gholami et al., 2016; Herwix et al., 2022; Majchrzak et al., 2016; Winter & Butler, 2011). This is demonstrated by several large-scale research efforts that have addressed real-world problems and used DSR frameworks to coordinate and integrate their work (e.g., Ginige et al., 2018; Nunamaker et al., 2013; Nunamaker et al., 2017; Paradice et al., 2019). For example, Nunamaker et al. (2013) developed and illustrated a DSR framework aimed at facilitating research programs that integrate all forms of IS research in pursuit of real-world impact. Ginige et al. (2018) described how they used a DSR framework to coordinate a complex, multidisciplinary research program to develop a solution to help farmers in Sri Lanka make better-informed decisions about which crops to plant. Paradice et al. (2019) illustrated how a grand challenge framework can be used to stimulate and coordinate a stream of research from nominally distinct groups, even over time frames of several decades.

However, despite these promising examples, this use of DSR is not yet part of the mainstream understanding of what DSR should be about. In particular, it would be difficult, if not impossible, to systematically publish work done as part of complex programs of research in top IS journals, simply because such work does not fit well into established DSR standards and templates (e.g., Gregor & Hevner, 2013).

Thus, more work is needed to establish DSR as a rigorous way to transcend the traditional boundaries of IS research and organize systematic research efforts that integrate insights and perspectives from non-academic practice, but also from other academic disciplines, including engineering, social sciences, and humanities, to address pressing societal problems (e.g., Briggs & Schwabe, 2011; Nunamaker et al., 2013; Nunamaker et al., 2017; Ram & Goes, 2021; Van de Ven, 2018; vom Brocke et al., 2021; Winter & Butler, 2011). In particular, a more flexible and open framework for DSR is called for, if we are to make its potential to act as a "super-methodology" for problem- and solution-oriented IS research in general accessible to mainstream IS research. The pragmatist understanding of science as merely a rigorous form of problem solving, provides an appropriate theoretical foundation for this conceptual work (cf. Morgan, 2007, 2014). Specifically, the work of John Dewey, as detailed in his *Logic: The Theory of Inquiry* (1938b), develops a compelling and useful account of how all forms of research can be seen as expressions of inquiry.

Three key concepts capture the essence of Dewey's theory (Brown, 2012). First, *situations* serve as both the source and control of inquiry, encompassing all relevant aspects of the environment (Brown, 2012, pp. 268–274). As such, a situation can be understood as the stage upon which inquiry unfolds. Second, *inquiry* is the interactive process of transforming indeterminate and problematic situations into determinate and resolved ones through purposeful engagement (Dewey, 1938b, pp. 104–105). The process of inquiry can be understood in terms of a linear sequence from the recognition of a situation as problematic to its resolution in the form of a final judgment, as well as nonlinear functional phases (e.g., institution of a problem, determination of a problem-solution, reasoning, experimentation) that guide the process (Brown, 2012, pp. 280–287; Dewey, 1938b, pp. 104–114). Finally, the *warranted assertibility* of the final judgment marks the legitimate endpoint of inquiry and provides a rigorous standard for knowledge generation. It is achieved only when rigorous experimentation has convincingly demonstrated that an indeterminate and problematic situation has been transformed into a determinate and resolved one (Brown, 2012, pp. 280, 297).

Overall, Dewey's theory anticipates key aspects of our mainstream understanding of DSR. It frames inquiry as a problem-solving process in which problematic situations are resolved through creative and purposeful engagement (cf. Hevner et al., 2004), emphasizes that all problem-solving can be understood in terms of general patterns (cf. W. Kuechler et al., 2005), and highlights the central role of experimental evaluation in determining scientific progress (cf. Baskerville et al., 2015). However, it goes beyond the mainstream understanding of DSR by recognizing that all forms of research, including BSR, can be understood in the same way because they are all fundamentally just different expressions of the broader process of inquiry. As such, Dewey's theory of inquiry provides a general, unifying framework that can facilitate the integration of diverse forms of research toward common goals.[5]

Second, this pressure on the mainstream understanding of the DSR paradigm is only intensified by the growing realization that in our increasingly turbulent digital age, where rapidly advancing IT developments are raising the stakes and even threatening the very fate of our collective game for the future of humanity (e.g., Center for AI Safety, 2023; Consilience Papers, 2022; Harari et al., 2023; Ord, 2020), we must now, more urgently than ever, come together to reflect on the *responsibility* this situation demands of us and act accordingly (Jonas, 1984). As IS researchers, we are part of a system of research and innovation that has not only positive but also increasingly recognized (unintended) negative

[5]See also Morgan (2014) for a similar argument in the field of social research.

impacts, such as the surveillance capitalism supported by data harvesting platforms (Zuboff, 2015), the threat to our epistemic commons posed by social media (Seger et al., 2020), or the perpetuation of discriminatory practices by black box algorithms (O'neil, 2016). A key question, then, is how to ensure that this system produces only *responsible* innovation.

Responsible innovation (RI) is an emerging concept and field of discourse that refers to the ideal of managing and governing emerging technologies (and potentially other innovations) in a way that is consistent with democratic values (Owen, 2019, p. 27). We focus on RI rather than other approaches to responsibility, such as corporate social responsibility or business ethics, because RI is unique in its focus on the societal dimension of responsibility. In a sense, it seeks to build on and integrate all other approaches and discourses of responsibility in the pursuit of sustainability and well-being at societal and global levels. Specifically, RI calls for innovation processes that are "more anticipatory, more reflexive, more inclusive, deliberative, open and, in total, more responsive" (Owen, 2019, p. 27), as well as the institutional and systemic changes needed to achieve this (e.g., Owen et al., 2021; Stahl, 2022). The goal is to establish "a collective commitment of care for the future through responsive stewardship of science and innovation in the present" (Owen, Stilgoe, et al., 2013, p. 36). This requires "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)" (von Schomberg, 2013, p. 63).

As such, RI challenges our modern understanding of innovation as the deliberate, human-made improvement of our circumstances, through the development and use of technology (Godin, 2014, pp. 2–4). It cautions us to recognize that innovation is not inherently good, but should be evaluated in terms of its consequences within a broader societal context (Stilgoe et al., 2013). It reminds us that innovation can even have catastrophic consequences if we end up losing control of our inventions (Genus & Stirling, 2018; Russell, 2019). At the same time, it recognizes that responsible forms of innovation are still urgently needed to rethink and remake entrenched but unsustainable systems and behaviors (e.g., Owen et al., 2021; Stahl, 2022).

That losing control is not an idle fear was already convincingly demonstrated by Collingridge (1980), who developed the *dilemma of control* to capture the insight that "attempting to control a technology is difficult, and not rarely impossible, because during its early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow" (p. 19).[6] To this day, finding ways to address this dilemma remains a key challenge for RI, inspiring calls for more open, participatory, diverse, incremental, flexible, and reversible governance processes that acknowledge our fallibility and ignorance in anticipating the societal consequences of emerging technological developments (Genus & Stirling, 2018).[7]

In this way, RI should not be seen as an answer or solution, but rather as a contextual question or challenge deeply rooted in a forward-looking notion of responsibility that leads us to ask: (1) what future we want to create; and (2) how we can responsibly bring about that desired future, given the inherent uncertainty and unpredictability of the future and

[6]Take the example of climate change. When we developed fossil fuel engines, we did not understand how the use of this technology would ultimately contribute to climate change. By the time we understood this relationship, the technology was already deeply embedded in society, making it very costly and difficult to reverse course.

[7]In a similar vein, *Jevons Paradox*—the recognition that efficiency gains in resource use tend to increase rather than decrease resource consumption (Giampietro & Mayumi, 2018)—is another example of how the outcomes of innovations can be counterintuitive and hard to fully anticipate (Blok & Lemmens, 2015).

our own ignorance and ambivalence about how to shape it (Owen, 2019, p. 29).

This understanding of RI calls for inclusive, inter-, multi-, and transdisciplinary collaboration and coordination on an unprecedented scale, with an unprecedented level of collective care aimed at shaping desirable futures that work for all (von Schomberg, 2019). As noted by Winter and Butler (2011), IS research seems well positioned to help answer such calls, as it has long been concerned with how to successfully design and implement socio-technical IS that could facilitate RI processes. For example, when RI calls for increased coordination through open research and innovation practices, this requires support from digital platforms and service ecosystems, and IS research has shown a strong interest in understanding how to successfully design, implement, and govern these (e.g., de Reuver et al., 2018; Hein et al., 2020; Sun & Gregor, 2023; Wang, 2021). RI also calls for more ethical methods of technology development, and IS research has a strong history of method engineering (Goldkuhl & Karlsson, 2020). In addition, IS research has at times demonstrated its ability to focus directly on addressing grand societal challenges in a responsible way (e.g., Davison et al., 2023; Majchrzak et al., 2016; vom Brocke et al., 2015; Walsham, 2017).

Ultimately, the growing demand for strategic, inter-, multi-, and transdisciplinary programs of research that address the grand societal challenges of our time through responsible innovation calls for a *paradigm shift* (cf. Kuhn, 2012) in IS research. We need to move beyond our current mainstream understanding of DSR and develop a *responsible DSR ecosystem* that can bring together all forms of research to address our common concerns and indeed ensure that the technologies we develop are appropriately controlled with the benefit of all affected parties and the whole system in mind (Ulrich, 1994; D. S. Wilson et al., 2023). In our technology-driven, interconnected, and globalized world, we cannot hide from this responsibility, but must embrace it (Jonas, 1984). The framework for responsible DSR proposed in this paper aims to contribute to the realization of this imperative.

2.2 Requirements for a Framework for Responsible Design Science Research

One of the key questions in any rigorous DSR effort is what the evaluation criteria should be for research results (Peffers et al., 2007; Vaishnavi et al., 2019; Venable et al., 2014). Here, we use the term *requirements* to refer to such evaluation criteria. Understood in this way, requirements can inform formative evaluations and thereby shape the direction of DSR efforts, but they can also guide summative evaluations that help assess the extent to which the final outcomes meet expectations (Venable et al., 2014). However, to the best of the author's knowledge, there is no generally accepted set of requirements for research frameworks such as the one presented in this paper. While the discourse in the philosophy of science is full of controversial discussions about the role of theoretical framworks in establishing research paradigms and contributing to scientific progress (Chalmers, 2013), it was not possible to find a compelling synthesis of the key requirements that such frameworks should ideally fulfill. Therefore, I decided to develop such a list myself by synthesizing key positions from the philosophy of science into a grounded set of relevant requirements for a framework for responsible DSR.

The starting point for this effort was the recognition that a framework for responsible DSR is usefully understood as a *boundary object* (Star & Griesemer, 1989). As Star and Griesemer (1989) pointed out, successful scientific work in practice requires not only competent individual action, but also systematic collaboration and, thus, mediated information exchange between different actors with diverse perspectives, motivations, and interests. They use the term *boundary object* to refer to the material means that are employed to facilitate this collaborative information exchange. In particular, "boundary objects are objects that are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use" (Star, 1989, p. 46).

While various types of boundary objects have been described (see Star, 1989), a framework for responsible DSR is best understood as an *ideal type*, which is an "object such as a map or atlas which in fact does not accurately describe the details of any one locality. It is abstracted from all domains, and may be fairly vague. However, it is adaptable to a local site precisely because it is fairly vague; it serves as a means of communicating and cooperating symbolically—a sufficient road map for all parties" (Star, 1989, p. 49). Understood in this way, a framework for responsible DSR seeks to provide an abstract description of responsible DSR practice that can act as bridge between different relevant perspectives, thus supporting the implementation of responsible DSR in different situations, as well as enabling collective efforts aimed at its further development. This understanding illustrates how frameworks such as the one developed in this paper, are a central means by which systematic research efforts are held together.

Given this understanding, one possible explanation for why there has been so little work on general requirements for such frameworks becomes clear. Because ideal-type boundary objects must appeal to the different and potentially competing perspectives of multiple parties, the question of their requirements is highly contextual and thus cannot be reliably answered in general terms (cf. Turnhout, 2009, p. 410). Our approach of synthesizing a set of context-specific requirements from the relevant literature is therefore an appropriate way forward.

For the purposes of this paper, Table 5 summarizes a set of seven relevant requirements for a framework for responsible DSR, which were chosen based on my engagement with key literature in philosophy of science and RI. For example, six requirements are derived from seminal texts in the philosophy of science and are complemented by a seventh requirement derived from the growing demand for responsible research and innovation, the relevance of which has also been recognized by global intergovernmental bodies (e.g., UNESCO, 2017). Together, they have been chosen to balance the requirements for theoretical robustness with practical applicability in complex, responsible DSR practice.

Table 5: Requirements for a Responsible DSR Framework.

Requirement	**Explanation**	**Sources**
Actionable	The framework should be actionable so that it can usefully inform DSR practice.	(Dewey, 1938b)
Comprehensive	The framework should be comprehensive in terms of covering key aspects of DSR practice, so that a rich program of research can emerge and a cumulative knowledge base can be developed.	(Kuhn, 2012; Lakatos, 1980)

Requirement	Explanation	Sources
Consilient	The framework should facilitate consilience with research perspectives from other fields and contribute to the growth of an integrated scientific knowledge base, so that a deeper and more useful understanding of the world can be achieved.	(Dewey, 1938b; E. O. Wilson, 1999)
Flexible	The framework should be flexible enough to be useful to the wide variety of DSR practices that already exist and are expected to emerge in the future, so that it remains useful over the long term.	(Kuhn, 2012; Lakatos, 1980; Popper, 2005)
Open	The framework should be open, accessible, and transparent to all relevant parties in order to facilitate productive inter-, multi- and transdisciplinary engagements and multifaceted critical discourses, so that effective solutions to even highly complex problems can be developed over time.	(Feyerabend, 1993; Gibbons et al., 1994; Popper, 2005)
Parsimonious	The framework should be parsimonious (i.e., as simple as possible, but as complex as necessary), in order to serve its purpose effectively and efficiently.	(Baker, 2022)
Responsible	The framework should help to strengthen the positive and limit the negative consequences of DSR practice.	(Douglas, 2009; Stilgoe et al., 2013; UNESCO, 2017)

In addition to being useful for the development and evaluation of the framework presented in this paper, the requirements can also be used to analyze the strengths and weaknesses of existing DSR frameworks in order to gain a better understanding of the current research landscape. Table 6 summarizes the strengths and weaknesses of four existing, more or less recognized DSR frameworks (e.g., Drechsler & Hevner, 2022; Hevner et al., 2004; Seidel & Watson, 2020; Vaishnavi et al., 2019) in addressing the above requirements.[8] I have chosen the four frameworks to be representative of the mainstream view of the DSR paradigm, so while other frameworks may exist, I argue that their analysis would not contradict the picture of the mainstream understanding of DSR that I paint here. In particular, my intention is to focus on frameworks that describe the mainstream understanding of the DSR paradigm in general, rather than investigating more specific frameworks that characterize different *genres* of DSR (Peffers et al., 2018). I argue that this focus is well justified given that the goal of this paper is to contribute to the development of an integrative, responsible DSR ecosystem that can bring together all genres of DSR, and even all forms of research in general, in pursuit of common interests.

My assessment shows that while all four DSR frameworks evaluated are strong in providing *actionable* guidance on some important aspects of DSR, only Vaishnavi et al. (2019) attempt to cover DSR *comprehensively*. However, while they mention the

[8] See the Online Appendix to this paper for a more comprehensive justification of our reasoning leading us to the final evaluation results summarized in Table 6.

ontological, epistemological, axiological, and methodological dimensions of DSR, they still fail to cover all dimensions in depth. The other frameworks limit themselves mainly to discussing the epistemological and methodological dimensions and only touch upon the other aspects in passing. Overall, the ontological and axiological dimensions of DSR practice are hardly considered in detail.

Table 6: Assessment of Existing DSR Frameworks.

Requirement	Hevner et al. (2004)	Vaishnavi et al. (2019)	Seidel and Watson (2020)	Drechsler and Hevner (2022)
Actionable	+	+	+	+
Comprehensive	±	+	±	±
Consilient	–	–	±	±
Flexible	±	±	+	+
Open	±	±	±	±
Parsimonious	±	±	±	±
Responsible	±	±	±	±

Legend: + Strength, ± Neither strength nor weakness, – Weakness

Furthermore, we suggest that none of the frameworks evaluated are strong advocates of *consilience* (E. O. Wilson, 1999), as they all assume that DSR is fundamentally different from other forms of research due to its interest in *prescriptive* rather than *explanatory* or *predictive* knowledge (e.g., Gregor & Hevner, 2013; Hevner et al., 2004). This perspective has severely limited our understanding of how DSR can be used and what contributions should be expected from it (cf. W. Kuechler & Vaishnavi, 2008; Levy & Hirschheim, 2012; Nunamaker et al., 2013). It does so unnecessarily because the logic of inquiry is similar no matter what outcome one is interested in (e.g., Dewey, 1938b; Farrell & Hooker, 2012; Hooker, 2017). There is no hard and fast rule that prohibits DSR from producing any and all types of knowledge (e.g., Baskerville et al., 2015; Dewey, 1938b). Simply put, rather than opening up broad horizons for DSR to contribute in any way that can be shown to be worthwhile, the frameworks prematurely narrow our view of what is possible by embedding rigid assumptions about what kinds of outputs DSR must produce. This understanding is at odds with the very idea of consilience, which seeks an integrated understanding of the world and, thus, advocates for open-mindedness about the boundaries to be drawn and open-endedness about the outcomes to be sought (Dewey, 1938b; E. O. Wilson, 1999).

In addition, this dichotomy also limits the *flexibility* of the frameworks by denying that some sound research efforts can be labeled as DSR. For example, a literal application of Hevner et al.'s (2004) framework would not allow novel research approaches that develop design knowledge based on observational studies, such as behavioral design science studies (e.g., Germonprez et al., 2011), to be identified as a form of DSR. Seidel and Watson (2020) and Drechsler and Hevner (2022) are more flexible in this regard, but still assume that DSR is inherently limited in the types of problem situations it can address and the outcomes it should produce. This limitation could be overcome if DSR was recognized as a "super-methodology" (Nunamaker et al., 1990) that can bring together all forms of research into strategic, inter-, multi-, and transdisciplinary programs of IS research (Nunamaker et al., 2013), rather than a research paradigm narrowly focused on the construction and evaluation of IT artifacts and prescriptions for their design, as was the case in all four frameworks evaluated.

With respect to all other requirements, the frameworks evaluated are neither particularly strong nor weak, with only minor differences among them. In particular, all of the frameworks are to some extent *open*, *parsimonious*, and *responsible*, but not remarkably so. For example, while all frameworks clearly aim to facilitate responsible rather than irresponsible behavior, there is almost no mention of the possibility of negative or unintended consequences from the research itself, or even a discussion of how to mitigate such risks. The growing need for better understanding and helpful guidance on responsible DSR makes this situation particularly problematic. In an era of accelerating technological progress and growing societal stakes, the need for responsible DSR is now more pronounced than ever. While some dedicated frameworks for responsible research already exist (e.g., Jirotka et al., 2017; Stilgoe et al., 2013), their insights have yet to be deeply integrated into the paradigmatic foundations of DSR.

3 Research Approach

This research project followed a multi-grounded theory development approach (see Figure 8). Multi-grounded theory development is characterized by the desire to achieve alignment in three key relationships (Goldkuhl & Cronholm, 2010): Between the empirical data and the emergent theory (i.e., empirical grounding); between the emergent theory and external theories (i.e., theoretical grounding); and between concepts within the emergent theory itself (i.e., internal grounding). For best results, Goldkuhl and Cronholm (2010) suggest that this process should be iterative and reflective, with an eye toward evolving research interests as new insights are generated.

In two major research iterations, I used empirical data generated through expert interviews as well as external theories identified in literature reviews to develop and explicitly ground an emergent theory of responsible DSR practice. In particular, the project began with a research interest in the opportunities and challenges for tool support in DSR (Research Iteration 1), but shifted to a stronger focus on the key challenges for responsible DSR (Research Iteration 2) when it became clear that a lack of shared understanding of DSR as a whole was a major bottleneck for any work aimed at supporting the development of a responsible DSR ecosystem. Since the interest of this paper is in the latter focus, I will mainly describe the work done in Research Iteration 2, and only sketch the work done in Research Iteration 1 for the sake of completeness and transparency.

3.1 Research Iteration 1

The first research iteration was initiated after a workshop at DESRIST 2017 revealed strong community interest in tool support for DSR (Morana et al., 2018). The goal was to complement the work done at the workshop with a more comprehensive, grounded perspective on tool support. To this end, a first round of 12 expert interviews was conducted between July and September 2018 to obtain more empirical data on researchers' experiences with DSR and how they would like to see DSR supported in the future (see the Online Appendix for more details). The interviews were semi-structured and sampled using a mixed logic of purposive and convenience sampling, which allowed us to integrate insights from different perspectives across the interviews, but also gave us space to explore the individual perspectives and interests of the interviewees. In addition, a systematic literature review on DSR was conducted to gain an understanding of existing theories on DSR as well as DSR tool support.

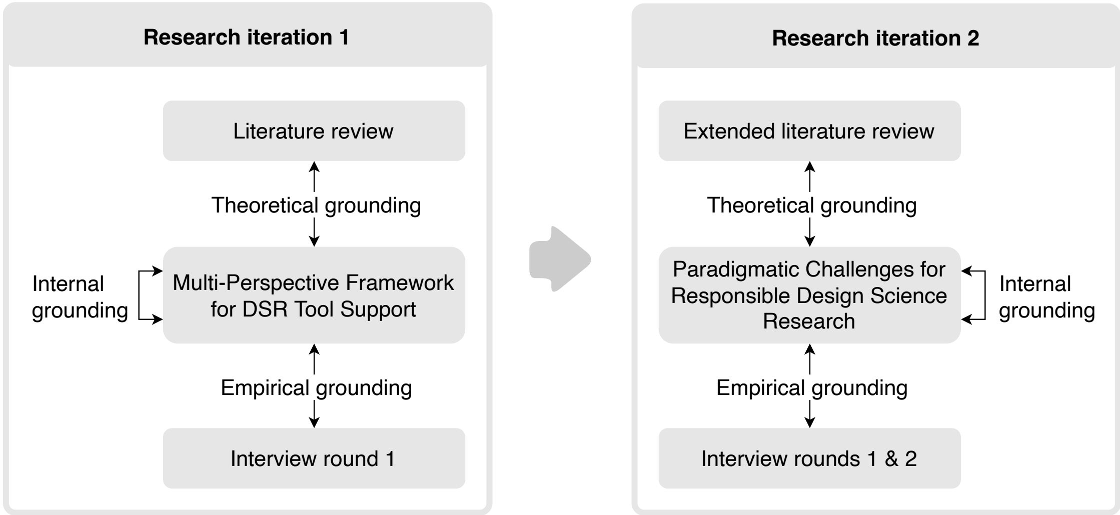

Figure 8: Multi-Grounded Theory Development Approach (informed by Goldkuhl & Cronholm, 2010, 2018).

Based on these data, we developed and tested a multi-perspective framework for DSR tool support, which organized a set of core stakeholder perspectives that we identified as useful to consider when thinking about the challenges and opportunities of DSR tool support (Herwix & Rosenkranz, 2019). However, as we reflected on our research interest, it became clear that a shift to a greater focus on responsible DSR in general was needed to make significant progress in DSR tool support. Although broad frameworks for DSR exist, they tend to severely limit our understanding of what DSR is and can be, and thus remain controversial (e.g., W. Kuechler & Vaishnavi, 2008; Levy & Hirschheim, 2012; McKay et al., 2012). Also, given the growing societal need for an integrative, *responsible* DSR ecosystem, I decided to shift the focus in the second research iteration to address this overarching issue more directly.

3.2 Research Iteration 2

Data Collection

The second research iteration built on the data and insights generated in the first iteration and expanded upon them with a broader, but still in-depth, examination of responsible DSR practice in general. I conducted an additional 29 expert interviews between April 2019 and October 2021. This second round of interviews had a stronger focus on uncovering interviewees' perspectives on what is and leads to high-quality DSR, as well as the opportunities and challenges associated with good DSR practice. Again, semi-structured interviews[9] were conducted to strike a balance between systematically covering relevant topics and exploring the interests and expertise of a particular interviewee. The participants in this round of interviews were sampled using a representativeness logic. Specifically, I contacted a set of 87 recognized IS scholars from around the world who were familiar with DSR, and also made two open calls for participation in the study on the AIS World mailing list in March 2020. Ultimately, I decided to combine the interviews from both rounds of interview into a comprehensive data set of 41 interviews. A summary of the

[9]See the Online Appendix of this paper for the updated interview guideline.

demographics of all of the participants (including the names of those who agreed to be publicly identified as participants in this study) and anonymized high-level summaries of the interviews can be found in the Online Appendix of this paper.

I also extended the literature review to how DSR is represented in the mainstream IS journals to get an even more comprehensive picture of how the IS research community thinks about DSR practice. To do this, I used the Web of Knowledge database[10] to repeatedly search the abstract, title and keywords of all papers through May 24, 2023 in the *Association for Information Systems Senior Scholars' List of Premier Journals*, *Business & Information Systems Engineering*, and the *Communications of the Association for Information Systems* as canonical sources of high-quality DSR. The keywords I used were: "design science" OR "science of design" OR "design research" OR "design-oriented research" OR DSR. This search yielded 527 articles, which were further screened for relevance by reading the abstracts and grouped into two categories for easier access and review: "Applied DSR" (386) and "Meta DSR" (141). In addition, I supplemented this systematic search with manual searches to identify reviews that addressed DSR in IS. In total, I found 91 reviews, which I grouped into five categories according to their objective (Paré et al., 2015): "summarization" (39), "data aggregation or integration" (0), "explanation building" (19), "critical assessment" (19), "hybrid" (14). The resulting reference lists can be found in the Online Appendix to this paper.

Taken together, these data sets provided a comprehensive basis for theory generation and explicit grounding of the emergent theoretical framework.

Theory Generation

I followed an iterative and flexible theory generation process that included *inductive coding*, *conceptual refinement*, *pattern coding*, and *theory condensation* steps, supported by the concurrent explicit grounding processes (Goldkuhl & Cronholm, 2010). Deliberate theorizing in the second research iteration began about halfway through the second round of interviews, when we began using the qualitative data analysis tool MAXQDA Plus 2022[11] to engage in the inductive coding of the opportunities and challenges mentioned in relation to responsible DSR. At this stage, the goal was to let "the data speak" in order to surface worthwhile themes that could be further refined and elaborated. Some codes at this stage were, for example, *"research that could be labeled as DSR is not"* or *"we don't know what is worth building"* under challenges, or *"resolve rigor vs. relevance"* or *"provide scaffolding and structure for research"* under opportunities. Through this process, relationships among the codes were identified and a hierarchy began to emerge around a core category ("Responsible DSR practice") with subcategories (i.e., key aspects of responsible DSR practice, such as "ontological perspective," "epistemological perspective," "axiological perspective," and "methodological perspective") that had multiple properties (e.g., "components," "examples," "opportunities," "challenges," "recommendations") specified in terms of dimensions (i.e., illustrative codes). I then proceeded to refine this coding hierarchy in many iterative and interwoven steps of conceptual refinement, pattern coding, and theory condensation. Several key decisions in this complex theorizing process were informed by the explicit grounding processes.

Explicit Grounding

A focus on three explicit grounding processes is the key differentiating factor between Goldkuhl and Cronholm's (2010) multi-grounded theory development approach and more

[10] https://apps.webofknowledge.com/

[11] https://www.maxqda.com/products/maxqda-plus

traditional forms of grounded theory where such processes are less clearly defined (e.g., Glaser & Strauss, 1967). The three explicit grounding processes are: empirical grounding, theoretical grounding, and internal grounding. Whereas empirical grounding aims to achieve alignment between the empirical data and the emergent theory, theoretical grounding aims to achieve alignment with external theories, and internal grounding aims to achieve coherence within the theory. Taken together, the grounding processes are proposed as a comprehensive support mechanism for the development of valid, acceptable, and understandable theories (Goldkuhl & Cronholm, 2010, 2018).

Explicit grounding shaped key decisions in the theorizing process. For example, in a flash of insight, the subcategories in the emerging coding hierarchy were understood to be deeply related to the key perspectives discussed in paradigmatic analysis, a well-established theoretical framework for analyzing research paradigms (e.g., Chen & Hirschheim, 2004; Iivari, 1991; Iivari, 2007). This has helped me to integrate and align our concepts more deeply with existing theories, while ensuring that the empirical grounding remained strong. Other examples of this kind of explicit grounding include the decision to align several of the key concepts with well-established scholarship, such as Ulrich's (2003; 2001, 2006, 2012a) work on critical systems thinking, Dewey's (1938b) work on the logic of inquiry, and Vervaeke and colleagues' work on problem solving and its relationship to relevance realization from a cognitive science perspective (Andersen et al., 2022; Vervaeke et al., 2012; Vervaeke & Ferraro, 2013b). These works have been highly influential in my theorizing, helping me to make better sense of the data and to move toward a more coherent alignment between the empirical data, internal theoretical constructs, and external theories.

In addition, the explicit grounding processes led me to develop a concept matrix to transparently document how consistently the interviewees had made at least implicit references to the key paradigmatic challenges I had identified. The results of this explicit grounding exercise can be found in the Online Appendix to this paper.

4 Four Paradigmatic Challenges for Responsible Design Science Research

Building on a critical pragmatist perspective, the multi-grounded framework conceptualizes responsible DSR as *inquiry* (see Figure 9). As Dewey (1938b, pp. 104–105) elaborates, inquiry can be understood as a knowledge-seeking process that is triggered by indeterminate situations that are perceived as *problematic* and is successfully concluded when these situations have been transformed to appear satisfactorily *resolved*. This lens is appropriate for framing responsible DSR because it is very much in line with the founding vision of DSR in the IS field: the development and evaluation of novel IT artifacts to resolve relevant problems in practice (e.g., Hevner et al., 2004; March & Smith, 1995).

Recognizing that a situation has become problematic is the first step in responsible DSR. This occurs when our established habits fail us in dealing satisfactorily with a situation (Dewey, 1938b, pp. 104–107). Thus, problematic situations can be characterized as *indeterminate* and *beyond current capacities*. They are indeterminate in the sense that it is unclear how the situation will unfold because there is no a priori correct way to frame or resolve them (Dewey, 1938b, pp. 105–108; Hayes et al., 1988). However, they can be domesticated and understood through processes of inquiry (Dewey, 1938b, pp. 104–114). Importantly, this does not mean that all problematic situations must be approached from a blank slate, or that all ways of framing a problematic situation are equally good and useful. Processes of inquiry build on the framings, knowledge, and norms that have proven to be useful in previous inquiries (Dewey, 1938b, pp. 12–14). In this perspective, situations become problematic when previous inquiries have not prepared us to handle them well—they remain beyond our current capacities. Further inquiry is needed to resolve such problematic situations.

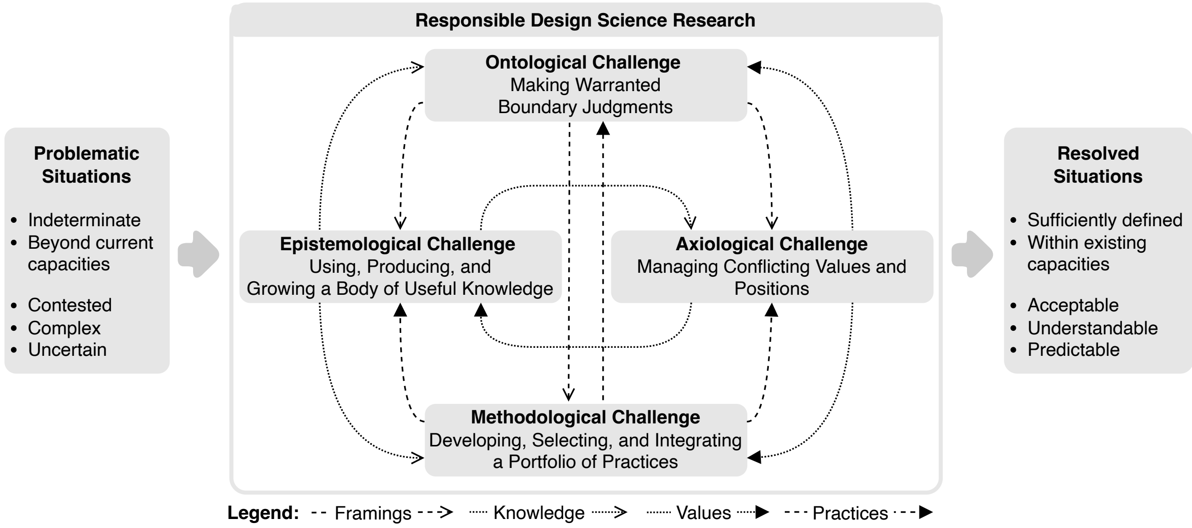

Figure 9: Four Paradigmatic Challenges for Design Science Research

In the context of responsible DSR, the problematic situations in focus are often characterized as *wicked problems* (e.g., Baskerville et al., 2018; Hevner et al., 2004; W. Kuechler & Vaishnavi, 2008). Wicked problems refer to social problems that do not have clear or correct solutions, in part because they are to some extent *contested, complex*, and *uncertain* (Rittel & Webber, 1973; Wanzenböck et al., 2020). The adjective *wicked* helps to emphasize that such social problems tend to have a mischievous quality, where proposed solutions turn out to be worse than the symptoms of the problem in the first place (Churchman, 1967). As such, they differ from narrowly defined natural science or technological problems that are "just" *complex* and tend to be solvable through traditional reductionist scientific study and mechanistic engineering (Rittel & Webber, 1973; Skaburskis, 2008). Because wicked problems are also more or less *contested* and *uncertain*, working on them tends to require more of an open, engaged, participatory, and transdisciplinary approach to inquiry that often involves crossing existing boundaries to engage relevant parties in the co-production of "clumsy" solutions that aim to address multiple conflicting values and positions (Hartmann, 2012; Wanzenböck et al., 2020).

Regardless of the actual approach, problematic situations become *resolved* to the extent that inquiry has transformed them in such a way that they can be dealt with satisfactorily as a whole (Dewey, 1938b, pp. 104–105). Given the nature of wicked problems, this means that, ideally, all affected parties consider the focal situations to have become sufficiently *acceptable* rather than contested, *understandable* rather than complex, and *predictable* rather than uncertain (Wanzenböck et al., 2020). Resolved situations can thus be distinguished from problematic situations in that they are *sufficiently defined* and manageable *within existing capacities*.

Against this background, the multi-grounded paradigmatic framework conceptualizes responsible DSR in terms of four interrelated *paradigmatic challenges*:

- *Ontological challenge:* Making warranted boundary judgments,
- *Epistemological challenge:* Using, producing, and growing a body of useful knowledge,
- *Axiological challenge:* Managing conflicting values and positions, and

- *Methodological challenge:* Developing, selecting, and integrating a portfolio of practices.

These challenges are *paradigmatic* in nature because they explicitly address the four major dimensions along which scientific paradigms tend to be characterized (e.g., Chen & Hirschheim, 2004; Iivari, 1991; Iivari, 2007; Morgan, 2007):

- *Ontology*: what are the nature and relations of being?
- *Epistemology*: what are the nature and grounds of knowledge?
- *Axiology*: what are the nature, types, and criteria of values and of value judgments?
- *Methodology*: what are the nature and body of methods in this particular field?

However, rather than characterizing responsible DSR in terms of specific answers to the questions posed by these dimensions, as has been done in previous paradigmatic analyses (e.g., Chen & Hirschheim, 2004; Iivari, 1991; Iivari, 2007), I want to clarify the nature of the questions: To what key challenges in responsible DSR practice do the ontological, epistemological, axiological, and methodological dimensions refer? This approach allows me to develop a framework that is not only *actionable*, but also *comprehensive* in addressing the key challenges of DSR practice. Moreover, it promotes *consilience* by providing an integrative perspective on DSR that is *flexible* enough to be applicable to the broad diversity of DSR (Rai, 2017a), as well as *open*, accessible, and transparent to relevant parties outside of the IS field. Yet, the framework remains *parsimonious* by focusing on the key challenges that are fundamental to *responsible* DSR practice.

The paradigmatic challenges are *interrelated* in that each of them depends on critical information produced by engaging with the others. For example, Ulrich (2003; 2000, 2006, 2012a) described how *boundary judgments* (the ontological challenge) depend on relevant *knowledge* (produced as a result of engaging with the epistemological challenge), *values* (negotiated as a result of engaging with the axiological challenge), and *practices* (selected as a result of engaging with the methodological challenge), while all of the other paradigmatic challenges in turn depend on the *framings* that are constructed through boundary judgments. Table 7 summarizes the four paradigmatic challenges and how they are interrelated with each other. The following sections discuss each of these challenges in more detail.

Table 7: The Four Paradigmatic Challenges for Responsible DSR.

Dimension	**Challenge**	**Definition**	**Information Flows**
Ontology	Making warranted boundary judgments	The ontological challenge emphasizes the need to be responsible when constructing the inherently selective framings that guide our inquiries.	Builds on relevant knowledge, values, and practices that are produced, negotiated, and selected through work on the other challenges. Constructs warranted framings that can direct the focus of responsible DSR inquiry.

Dimension	Challenge	Definition	Information Flows
Epistemology	Using, producing, and growing a body of useful knowledge	The epistemological challenge refers to the need to orchestrate the use and production of knowledge-as-object and knowledge-as-action when working toward resolving problematic situations and growing a useful body of knowledge.	Builds on relevant framings, values, and practices that are constructed, negotiated, and selected through work on the other challenges. Produces useful knowledge in the form of knowledge-as-object and knowledge-as-action that can support and advance the work on all other challenges.
Axiology	Managing conflicting values and positions	The axiological challenge highlights the need to find strategies and mechanisms for facilitating acceptance and agreement between competing stakeholder values and positions.	Builds on relevant framings, knowledge, and practices that are constructed, produced, and selected through work on the other challenges. Negotiates values and positions that can inform and shape all inquiry-related evaluative judgments.
Methodology	Developing, selecting, and integrating a portfolio of practices	The methodological challenge focuses on the need to develop, select, and integrate a portfolio of practices that satisfy the specific requirements of the problematic situation at hand.	Builds on relevant framings, knowledge, and values that are constructed, produced, and negotiated through work on the other challenges. Selects a portfolio of practices that help to steer the overarching responsible DSR process and carry out the work on all other challenges.

4.1 Ontological Challenge

Making Warranted Boundary Judgments

In a complex and entangled world like ours, actors with limited resources and capacities like ourselves must make decisions about where to focus our attention and energies (e.g., Vervaeke et al., 2012; Vervaeke & Ferraro, 2013b). Doing this responsibly requires us to make *warranted boundary judgments* that help us focus on the parts of the world that matter in a given situation (Dewey, 1938b; Ulrich, 2001, 2006, 2012a, 2012b; Ulrich & Reynolds, 2010).

As visualized in Figure 10, for the purposes of responsible DSR, this means that we must develop and ensure *logical coherence* between the *empirical* and *theoretical framings* that bound and guide our inquiry (Dewey, 1938b, pp. 108–114), and subject the inherent *selectivity* in our boundary judgments to ongoing *critique* and reflection (Ulrich, 2001, 2006, 2012a; Ulrich & Reynolds, 2010). This dual aspiration is grounded in the understanding that while we can use coherence as a criterion for evaluating the logic that connects empirical experience and theoretical framing (Dewey, 1938b, pp. 104–114), we must also acknowledge that we are always selective in what we focus on and ensure that this selectivity does not cause unnecessary or unintended harm (Ulrich, 2006; Ulrich & Reynolds, 2010). As Ulrich (2001, 2006, 2012a) has eloquently argued, while this problem of selectivity can never be fully resolved, it can be iteratively and productively addressed in practice. This requires making the selectivity of our framings explicit, so that our choices can be critically discussed. The goal is an emancipatory discourse that brings all concerned parties closer to collective agreements on how to frame given situations. But even if collective agreements remain out of reach, the very act of enabling critique can help those affected gain more recognition for their needs. Ultimately, the best that can be hoped for are at least critically tenable framings of problematic situations.

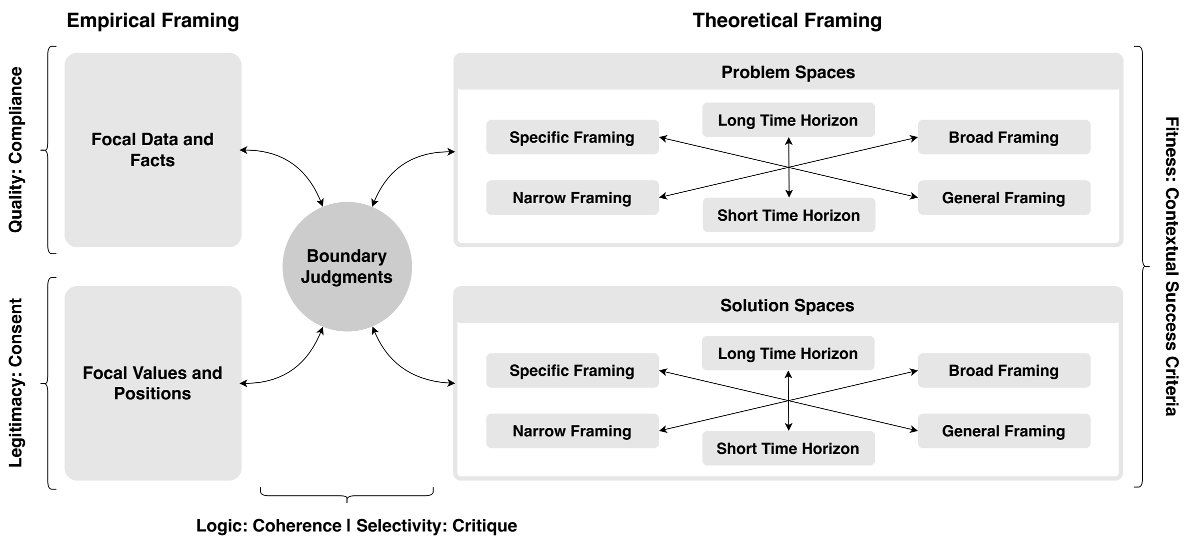

Figure 10: Boundary Judgments in Responsible Design Science Research.

In this view, an empirical framing is developed by making boundary judgments about the *data and facts*, as well as the *values and positions*, that will form the empirical basis of the inquiry. Decisions about what to include and what to exclude are highly interdependent, but also related to decisions about the theoretical framing—the *problem* and *solution spaces* that are constructed as part of a particular inquiry (Simon, 1973). For example, in a discussion of a real-world DSR project, one interviewee emphasized how adding additional values and positions to the process of developing a mobile health app shifted the overall conception of the problem, and, thus, how they framed the project's problem and solution space:

> Our goal [. . .] was to improve the operational efficiency of physicians. [. . .] But when we first showed the first version of the app to physicians, they said: "This is great, but we don't think the patients are going to use it." [. . .] We

> actually made the effort to go talk to patients, which was really difficult. [...] And we found that patients really liked the app, and they contradicted a lot of what physicians had told us patients wanted. [...] So as we thought through this and as we went into the next version of the app, we shifted our measure of success. [...] Now we're talking about service-dominant logic and the idea that in a service operation, there's value co-creation between the players. In this case, it's a physician and a patient. And when you create a product like that, if you focus only on efficiency, the whole service operation suffers. (I-2-2)

The same example also illustrates the importance of assessing the *quality* of focal data and facts in terms of their *compliance* with established quality criteria, such as their completeness in covering relevant aspects of the phenomenon (Batini et al., 2009). In addition, the *legitimacy* of the entire empirical framing was called into question because not all affected parties had a voice and expressed *consent* with the goals and outcomes of the inquiry (see Ulrich & Reynolds, 2010). Although at first glance physicians might be seen as appropriate proxies for assessing patients' goals and needs, this was not the case:

> So physicians had told us: "Be careful with trends," for example, "because the trends for a cancer patient are only going down. They're only going to get worse and that's going to be depressing. Show us the trends, don't show them to the patient. Be careful with the terms, don't make the terms too difficult." Well, the patients contradicted that. They're like: "We know our trends are going down, but it makes us feel better to see that. We can reassure the people who take care of us, we want to see them, we want all the information we can get." And then about the terms, they said: "We know every single term because the minute we are diagnosed with this disease, we go on the Internet. We find out everything there is to know, and we take notes during our visits. So it's also insulting that they would think that we don't know these terms." (I-2-2)

This example clearly demonstrates that ongoing reflection and critique of the selectivity of boundary judgments about whose voices are to be heard is paramount in responsible DSR. Without a conscious effort to include all affected parties, we risk making avoidable misjudgments that contribute to perpetuating rather than resolving problematic situations.[12]

Similarly, boundary judgments about the scope of the problem and solution space should also be subject to ongoing critical reflection. In particular, decisions about the *level of abstraction* (specific to general), the *logical scope* (narrow to broad), and the *time horizons* (short to long) over which problem situations and potential solutions are framed will always need to be made and should be critically reflected upon (Baskerville et al., 2015; Porra, 2001; van Gigch et al., 1997).

In terms of level of abstraction, Baskerville et al. (2015) illustrate how DSR inquiry can generate knowledge and insights across the spectrum from the idiographic (i.e., specific and situational) to the nomothetic (i.e., general and abstract). However, researchers' attention and resources are limited, so decisions about which levels of abstraction are most

[12] Where the affected parties cannot be consulted directly, as in the case of future generations or non-human animals, it is necessary to find or develop appropriate proxies that help us to at least approximate their values and positions. For example, in a recent treatise on animal justice, Nussbaum (Nussbaum, 2023) describes how it is quite possible to use well-designed experiments to gain a good understanding of the needs of nonhuman animals and to provide a reasonable basis of information that can inform our behavior toward them.

appropriate to focus on still need to be made and reevaluated over time. As vom Brocke et al. (2020) argue, DSR can be understood as a journey through space and time that depends on the evolving needs and demands of the world.

In terms of logical scope and time horizons, the work of C. West Churchman (van Gigch et al., 1997; see Porra, 2001; Ulrich, 2004) and other systems theorists (e.g., Ackoff, 1994; Reynolds & Holwell, 2010) as well as sustainability-oriented researchers (e.g., Loorbach et al., 2017; Schlaile & Urmetzer, 2019) emphasize the ethical imperative not to view problems and solutions too narrowly and only over short time horizons, but also to consider them in relation to the whole system (i.e., the broader context) over longer time horizons, because all human activities are now more or less deeply interconnected on a global scale. Ultimately, if we are to continue to thrive and survive, we must learn to make decisions with the well-being of the whole Earth system and our future generations in mind (Consilience Papers, 2022; Jonas, 1984; Ord, 2020; Ulrich, 1994; D. S. Wilson et al., 2023). This recognition is gaining traction in the ongoing debates about the sustainability, side effects, and dark side of digital technologies:

> The issues of sustainability, etc. [...] will become more and more important. [...] People develop algorithms and they evaluate the performance of algorithms... That's been done for a long time... But they also say that it's important to have algorithms that don't use too many resources, because that's bad for the environment, right? In terms of computing power, storage, etc.. So, I think that the future is really bright for design science research, and at the same time we will probably take a more holistic approach and start considering [...] the side effects of our artifacts and other issues. [...] They sometimes take a long time to materialize, but it's a very important issue, so we have a responsibility to consider the different impacts of the artifacts that we produce. (I-2-14)

However, what exactly qualifies as sustainable, ethical, or responsible in any given situation is generally a matter of context that cannot be fully anticipated in advance (Andersen et al., 2022; Dewey, 1938b). Ongoing evaluation is therefore necessary to ensure that what one is doing remains appropriate to the needs of the situation. In particular, this involves critically reflecting on theoretical framings in terms of their *fitness* for purpose and their ability to meet *contextual success criteria*. This is a pragmatic approach that is consistent with how we develop in any professional practice (Dewey, 1938b; Schön, 1984). It is well illustrated by the example of mobile health app development mentioned above. As the interviewee pointed out, the theoretical framing of the problem situation and potential solutions shifted significantly during the research process, otherwise *success* would not have been achieved:

> We're not going to get operational efficiency if the patients don't use the app, and we're not going to get the patients to use the app unless we look at this as a service. And the fact that in order to do a service, you have to involve both the customer and the provider. And then it becomes more of a service-dominant logic viewpoint. So I think that's the hard part of design science. You're solving real problems, and when you see those problems in the field, the measure of success changes [...]. (I-2-2)

Overall, the ontological challenge focuses on the practical need to identify what is justifiably and legitimately relevant to a particular inquiry. In the context of responsible DSR, this means making warranted boundary judgments about the empirical and theoretical framing of an inquiry. To be warranted, boundary judgments must be evaluated for logical coherence and sufficiently critiqued for their inherent selectivity. Additional criteria to be evaluated are the compliance of relevant data and facts with established quality standards, the involvement and subsequent legitimating consent of all affected parties, and the fitness for purpose of theoretical framings assessed in terms of contextual success criteria.

4.2 Epistemological Challenge

Using, Producing, and Growing a Body of Useful Knowledge

A key goal of responsible DSR is to go beyond resolving problematic situations through trial and error by identifying, documenting, and sharing key insights and lessons that help us resolve problems more productively over time (e.g., Baskerville et al., 2018; Gregor & Hevner, 2013; Hevner et al., 2004; vom Brocke et al., 2020). I call this the challenge of using, producing and growing a body of useful knowledge. Inspired by Cook and Brown (Cook & Brown, 1999), this challenge is framed as a quest for a generative dance between *knowledge-as-object* manifested in *boundary objects* (Star, 1989, 2010; Star & Griesemer, 1989) and *knowledge-as-action* expressed in our personal *ways of knowing* (Vervaeke & Ferraro, 2013b; Vervaeke & Mastropietro, 2021) in the practice of *inquiry* (see Figure 11).

The generative dance is guided by the open-ended rhythm of inquiry, which moves iteratively through four general stages: *framing* the situation, *planning* a strategy, *implementing* action, and *evaluating* the results (Dewey, 1938b, pp. 104–114; see also the generic design cycle in Vaishnavi and Kuechler, 2015). It is not successfully completed until the dance has transformed the problematic situation in focus so that it is appropriately resolved, ideally forming a unified whole (Dewey, 1938b, pp. 104–105). Thus, knowledge-as-object (i.e., boundary objects) and knowledge-as-action (i.e., ways of knowing) are not only the participants but also the products of this dance. The more their interplay is recognized and understood, the more generative and responsible their dance can become (Cook & Brown, 1999).

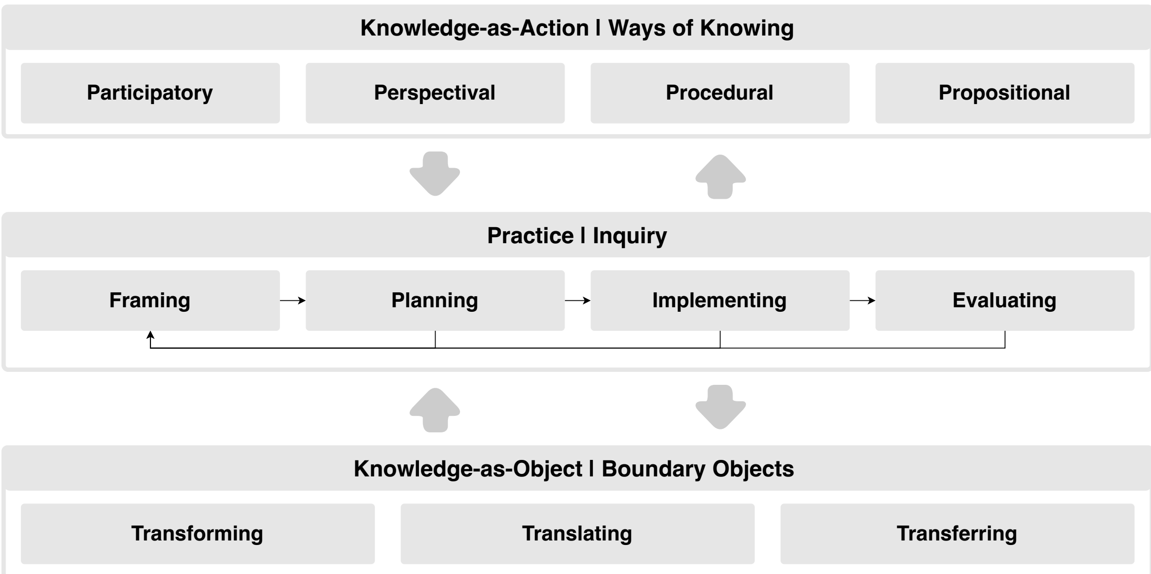

Figure 11: Inquiry as a Generative Dance Between Knowledge-as-Object and Knowledge-as-Action.

On the one hand, knowledge-as-object emphasizes the instrumental nature of knowledge, framing it as a thing or an artifact that can be "possessed" and used to achieve some end

(Cook & Brown, 1999). I conceptualize knowledge-as-object as boundary objects because this draws attention to the social dimension of knowledge use, production, and exchange in practice (e.g., Carlile, 2002, 2004; Fleischmann, 2006; Pennington, 2008; Star & Griesemer, 1989). Boundary objects are a central means by which effective coordination among actors can be achieved (e.g., Star & Griesemer, 1989). They are central to inquiring activities, such as new product development or scientific research, where multiple parties interact to challenge existing practices and develop new ways of framing, knowing, organizing, and doing (Avelino et al., 2019; Carlile, 2002, 2004; Fleischmann, 2006; Pennington, 2008). Thus, I join others in arguing that the effective and efficient creation, use, and adaptation of boundary objects in social practice is what drives inquiry toward the resolution of problematic situations and thereby generates knowledge (e.g., Carlile, 2002, 2004; Fleischmann, 2006; Pennington, 2008). Following Carlile (2004), I highlight three ways in which boundary objects can be used to use, produce, and grow a body of useful knowledge in responsible DSR practice:

- Boundary objects support knowledge *transformation* when a situation has become problematic (e.g., existing circumstances do not meet expectations) and new (political) arrangements need to be negotiated. This involves the use of visual inquiry tools, prototyping, and other types of boundary objects that facilitate the exploration, negotiation, and establishment of common interests and (political) arrangements.

- Boundary objects facilitate knowledge *translation* when common meanings need to be developed among relevant parties. For example, boundary objects such as concept maps can be used to facilitate cross-functional, inter-, multi-, or transdisciplinary interactions that occur as part of planning or implementation activities and lead to shared meanings.

- Boundary objects enable knowledge *transfer* when common meanings and interests can already be relied upon, so that it may be sufficient to store, copy, or move data via boundary objects to facilitate productive interactions. To this end, a context-specific common lexicon must be developed to help relevant parties to share and access the knowledge in question. Standardized user interfaces, standardized forms, or scientific reports are examples of boundary objects that can facilitate such knowledge transfer.

Knowledge-as-action, on the other hand, emphasizes the dynamic and enacted nature of knowledge, framing it as an action that is "practiced" or, more precisely, recursively constituting and constituted by practice (Cook & Brown, 1999; Orlikowski, 2002). This view complements our understanding of knowledge-as-object by emphasizing that the notion of knowledge only makes sense in relation to practice (Cook & Brown, 1999; Dewey, 1941; Orlikowski, 2002). I conceptualize knowledge-as-action in terms of ways of knowing because this allows us to consider the key dimensions of cognition that enable the emergence of intelligent (i.e., the ability to learn), rational (i.e., the ability to learn learning), and wise (i.e., the ability to harmoniously balance all relevant aspects of behavior) behaviors and practices (Vervaeke & Ferraro, 2013b; Vervaeke & Mastropietro, 2021). Following Vervaeke and Mastropietro (2021), I highlight four ways of knowing as particularly relevant for facilitating inquiry (see Table 11.1 in Vervaeke & Mastropietro, 2021):

- *Participatory knowing* is a fundamental form of knowledge-as-action that refers to an agent's ability to realize relevant ways of life (existential moding). It provides motivation to participate in the arena of life, which may culminate in a sense of *religio*[13] felt in terms of *resonance* (e.g., we feel resonant with our way of life).

- *Perspectival knowing* is at the core of our subjective phenomenology as agents in an arena. It refers to an agent's ability to realize relevant (i.e., fitness-enhancing) states of consciousness and cognition (salience landscaping). It provides perspective, is extended through *roles*, and can be perceived in terms of *presence* (e.g., something is salient to us in a situation).
- *Procedural knowing* refers to skills, habits, and the ability for perform relevant interactions (sensorimotor interaction). It is deeply intertwined with the creation and use of technologies that require, but also help to leverage, procedural knowing. It provides competence, is embedded in *routines*, and can marshal *power* (e.g., we are empowered to do something).
- *Propositional knowing* captures our ability to realize relevant implications (language and inference). It has probably become the most revered way of knowing in our societies. It provides reason, is embodied in *rules*, and produces *conviction* (e.g., we hold certain beliefs).

The relevance of these four ways of knowing (i.e., knowledge-as-action) becomes clear when considering the use of IT artifacts (i.e., knowledge-as-object) in practice. As Pentland and Feldman (2008) illustrate, IT artifacts tend to harness and support propositional knowing by enabling engagement with useful information that would otherwise be difficult and costly to maintain. However, as they eloquently point out, IT use in practice is an emergent and highly contextual phenomenon. To understand it, one must look closely at the parties involved and examine their patterns of action. In particular, the use of an IT artifact depends on the user's ability to perform the necessary interactions with the IT artifact (i.e., procedural knowing), the user's understanding of the structure and purpose of the interactions (i.e., perspectival knowing), and the user's worldviews, attitudes, and background motivations for using the IT artifact (i.e., participatory knowing). Thus, realizing positive value from IT artifacts requires not only the design of useful artifacts (i.e., knowledge-as-object), but also the fostering of synergistic ways of knowing in their users and other affected parties (i.e., knowledge-as-action). This perspective is also vividly illustrated by one of the interviewees who discussed the knowledge development potential of Action Design Research (ADR) approaches (Sein et al., 2011) in a developing country context:

> So most of the time what happens is that the NGO's, the organizations from the Western world, they just come with some tools and then they hand them over. And after a year the tools actually disappear. No capacity building, no other better world. It's just a waste of time. And I have seen it. The reason is because whoever goes there with this technology, they don't try to understand the context, they don't work with the local champions. That's the problem. So a method like ADR is very much about working with local champions who actually know the context. And then also from our perspective, as I said, we need this abstract knowledge, the principles. From their perspective, the practical use. What kind of affordances do they see or don't they see, and what kind of information can we provide so that they can see those things.

[13]The Latin term religio is used in this context to point to "a sacral experience of connectedness and transpersonal relation to reality" (Vervaeke & Mastropietro, 2021, paragraph before Table 11.1).

> So it's a very close proximity between the local champions and the academic researchers. That's the beauty of this area, if it's done right. (I-2-19)

Taken together, the interplay between knowledge-as-object and knowledge-as-action creates indeterminate situations with room for *virtuous cycles* of rapid knowledge production and associated growth in actors' wisdom (i.e., their ability to synergistically deploy, grow, and deepen ways of knowing) in using that knowledge, as well as *vicious cycles* of reinforcing self-deception and knowledge degradation or loss, and everything in between (Vervaeke et al., 2012; Vervaeke & Ferraro, 2013b).

These relationships are well illustrated by an interviewee who elaborated on this issue in the context of tool making in scientific practice. On the one hand, there are opportunities for virtuous cycles based on synergistic effects between powerful tools (i.e., knowledge-as-object) and improved capacities to use these tools in scientific practice (i.e., knowledge-as-action):

> Coming from computer science, [. . .] nobody sits down and writes software if they're not trying to solve some problem, and that's a tool. Computers are tools, they're meta tools in a way, that let us build tools, that let us run tools. There's a whole stack, we call it a stack of tools that help you build tools. [. . .] The more you get into the mindset of [. . .] being able to understand different tools and [. . .] what kinds of problems those tools solve, the better you are as a professional, the better you are at solving problems in the future. [. . . But] as long as individual researchers are expected to solve problems with just tools that already exist and that they can learn quickly, they're going to be very limited in the problems they can address. We would not have the ability to go to the moon, the ability to read genomes. Many of the things that science does depend on large investments in tools. [. . .] To the extent that research is done [. . .] only with the tools that already exist, to that extent [. . . it is] very limited in the problems [. . . it] can address. (I-2-1)

On the other hand, the interviewee also emphasized the risk of vicious cycles leading to loss of knowledge and even personal despair if we delude ourselves about our ability to effectively use the tools on which we depend:

> I had a relative who was doing a doctorate [. . .] back in the 1970s, and he was keeping all of his research data on a two and a half megabyte disk platter. [...] He had worked for several years collecting this data, all the data was on this disk platter. [One day,] he stuck his disk in the drive. [. . .] There was a head crash. It wiped out his data, and he was screwed. He gave up. He actually had a mental break down. It was really serious. He didn't understand that the tool he was using wasn't a perfect repository of data. It needed to be backed up, and he had not backed it up. So he had a very powerful tool, but he didn't understand how to use it. He didn't understand the infrastructure that was part of using it correctly. It's no different from a home owner who needs to cut down a tree and goes out and buys a chainsaw and ends up cutting his foot off because he doesn't know how to use it. (I-2-1)

Worryingly, such risks of vicious cycles are not limited to the personal sphere but also exist in the societal sphere. The inability of our societies to use advanced technologies wisely and the resulting potentially catastrophic risks have become a common research topic, for example, in research on sustainability transitions (e.g., Loorbach et al., 2017) or global catastrophic risks (e.g., Bostrom & Cirkovic, 2011). As a result, *responsibility* has been advocated as a key value and principle for innovation and research policy (e.g., Jirotka et al., 2017; UNESCO, 2017; von Schomberg & Hankins, 2019). In particular, it is increasingly accepted that we need to actively manage the risks associated with our research efforts in order to increase the chances of virtuous cycles and minimize the risks of vicious ones (e.g., Consilience Papers, 2022; Jirotka et al., 2017; Muehlhauser & Salamon, 2012; Stahl, 2022). This requires a productive and competent interplay between knowledge-as-object and knowledge-as-action to be constantly adapted and renegotiated, much as a successful dance between two dancers depends on careful ongoing coordination.

An example that illustrates this challenge close to home is the question of how to orchestrate this interplay in the context of education. As one interviewee pointed out, the fruits of our work also need to be translated into teachable formats (i.e., knowledge-as-object) so that they can actually be used in practice (i.e., knowledge-as-action), but how to grow the integrated body of knowledge that would be necessary to organize this in an efficient yet responsible manner remains an open question for the DSR community:

> The question is, how do we get the scientific knowledge into our curricula so that it is used? I think that's largely influenced by the confidence that we have in that knowledge. You know, if I publish a paper in EJIS and we've done a case study and we've done a fair evaluation, but we don't have a hundred or a thousand cases where people have used it, like a million bridges have been built out there. So why would you actually include this particular paper in your curriculum? There are some examples of design principles that are actually taught in practice. Think about object-oriented software development for instance, right? That is probably the most famous, but there are others. The question is, what do we need to do in our research, and how do we need to justify and test what we are doing, in order to get it to a state where it finds its way into our curricula? Then people will use it. And, this is my last comment on that: To what extent can we actually do that? Think about technological change and digital technologies and what's going on there. Think about the level of abstraction that would need to be required in order to do that. I think that's another challenge for the DSR field. (I-1-10)

In summary, the epistemological challenge highlights the need to find ways to orchestrate a responsible interplay between knowledge-as-object and knowledge-as-action in responsible DSR, so that the chances of virtuous cycles are enhanced and the risks of vicious cycles are minimized for all concerned.

4.3 Axiological Challenge

Managing Conflicting Values and Positions

Managing conflicting values and positions is at the heart of all forms of inquiry, as trade-offs between competing concerns are inevitable (e.g., Andersen et al., 2022; Simon, 1972; Tatar, 2007; Ulrich, 2003; Ulrich, 2006; Wanzenböck et al., 2020). As one interviewee pointed out, this connection is so deep that acknowledging this challenge may even be the central path to scientific insight:

> Design conflicts are the issue. This is the thing to go for when you really want to investigate the design of systems in order to develop scientific insights. Because if there are no conflicts, I mean, come on. There might still be problems. It

might be difficult to solve a certain problem, maybe even impossible, but as soon as you start dealing with conflicts where [...] you have two goals and both of them make perfect sense. And then you realize: "Oh, yeah. They make perfect sense, but they're in conflict. So what can I do?" It doesn't always work, but sometimes you can relax the conflict. There are a few strategies that help you do that, and that is how you develop new knowledge. (I-2-15)

But even if one is less enthusiastic about the role of value conflicts as a doorway to scientific progress, they naturally come to the fore for all of us when inquiry stalls because of them. This is often the case when problem framings or proposed solutions become highly contested (Wanzenböck et al., 2020). Due to the interdisciplinary and transformative nature of responsible DSR and its focus on wicked problems, we may encounter such conflicts at all stages of inquiry (Herwix et al., 2022). For example, a recurring theme in the interviews was that researchers face competing interests between working to publish contributions to academic discourse and helping practitioners and societal stakeholders improve their own conditions:

> Most people [...] who work on DSR have no clue about real-world problems and are not interested. But they are interested in getting their papers published, and I think that is really bad. I think there should be much more of a commitment in the whole community to collect real-world problems and [...] have a very timely discourse about innovative solutions to tackle them. (I-1-12)
>
> There are some influential people in our community who proclaim just one simple goal: "Get your work into A journals." How do you do that? It's not about results. It's about publication and selection in A journals. [...] When I see how many highly talented, motivated, young people are attending courses on how to do A journal contributions. This is terrible. It is really terrible. It is pure opportunism. It doesn't have anything to do with what can be counted as results, with benefit to society. It's all about benefit in terms of publication impact and h-factor. (I-2-18)

A key question in this context is how we can align our academic aspirations and incentives to produce rigorous and novel knowledge with the needs and demands of stakeholders outside of academia:

> Gradually, I think, the tax community is asking: "What do we get for the [... tax] dollars being spent [... on research]? Show us outcomes." And the research community has not been able to do that in any significant way compared to the investment that goes into it. So, they [(i.e., governments)] are changing the whole way they evaluate research, and I hope that would push people to think more about the practical outcomes that will help the community that is paying for our research. (I-1-2)

However, conflicting values and positions can exist not only between academia and its external stakeholders, but also within academia itself. For example, how to evaluate the quality of each other's work is a key issue on which any academic community must reach some agreement:

> I think the main problem when it comes to publishing right now is that there is a huge disagreement between different scholars about what good DSR is. I mean, it is very common, both at conferences and in journals, to get completely

> opposite reviews. I don't know if you've experienced this, but it can happen that you get one that's super, super, positive and then another that's very negative with a completely different view of how things should be done. (I-2-9)

These examples illustrate the contested nature of DSR at all levels. They point to the need for strategies to help us manage conflicting values and positions so that research can remain productive and beneficial to all concerned. Finding the right strategies can be challenging, however, because there is no master plan for agreeing on which values and positions to prioritize in the event of conflict. Each perspective is inherently selective and incomplete, which means that there is no fully rational solution to this problem (Ulrich, 2006; Vervaeke et al., 2012). As finite beings with limited perspectives, all we can do is to identify what is relevant to us and others and negotiate trade-offs between competing values and positions that we find both individually and collectively acceptable (e.g., Farrell & Hooker, 2013; Vervaeke et al., 2012; Wanzenböck et al., 2020). Yet we need not despair, for our success as a cooperative species shows that we are quite capable of doing so effectively under certain conditions (e.g., Ostrom, 2010; D. S. Wilson et al., 2013).

Based on my analysis of the interviews, I suggest that sustainable strategies for managing conflicting values and positions can be usefully characterized in terms of how they blend different *degrees of control* (see Figure 12). At one end of the spectrum, they tend to have *low-control strategy components* that create spaces for an open-ended exploration of different values and positions to address creative design challenges:

> I think the worst thing we can do is to restrict the creativity of a designer saying how it should be done. (I-1-2)

At the other end of the spectrum, they may incorporate *high-control strategy components* that use rules, regulations, and technology to routinize, standardize, and even automate individual behaviors in order to coordinate behaviors, ensure the quality of processes and outcomes, and increase the efficiency of stakeholder interactions:

> There needs to be some form of standard to document knowledge in order to be able to integrate it and compare it. So yes, there should be some... I don't know if it should really be a standard or some suggestions for structuring [knowledge] but then if there is some [standard] and everyone or most people somehow stick to it, it would be easier to integrate knowledge. (I-1-3)

The two strategic poles address critical concerns that are all needed at certain points in the resolution of problematic situations (Andersen et al., 2022; Vervaeke et al., 2012; Vervaeke & Ferraro, 2013a). They should thus be seen as complementary. In this sense, sustainable strategies for managing conflicting values and positions require an appropriate and dynamic mix of low-control and high-control strategy components. As one interviewee put it:

> In the design process, we need to work out: What are things that are a bit more mechanical and what are things that are a bit more individual or creative? (I-2-23).

A particularly useful way to facilitate sustainable management is to combine low-control strategy components to address challenges that have a strong creative dimension and benefit from exploring diverse perspectives, with high-control strategy components to focus resources where there is already agreement that it is of benefit to all concerned (e.g., Seidel et al., 2010). For example, academia emphasizes the academic freedom of individual researchers with the goal of harnessing the specialized expertise and passion of researchers to encourage a creative exploration of diverse topics (i.e., a low-control strategy component). At the same time, institutionalized peer review (i.e., a high-control strategy component) focuses substantial amounts of academic resources into a general system that controls the academic discourse and directs it toward good practices by exploiting the pre-existing knowledge embodied in the expert opinions of the peer reviewers.

Figure 12: Strategies for Managing Conflicting Values and Positions.

Such mixed strategies can be very successful in managing conflicting values and positions because, on the one hand, real-world interactions involving people are often highly complex in ways that quickly exceed the requisite variety of high-control strategy components. Low-control strategy components are needed to leverage specialized expertise to explore the broader terrain by motivating individuals to engage the subject-matter on their own terms:

> Well, the first reason [for doing research] should be... Because we like it, we are fascinated by it, and we hope it will help us learn more about the things we are interested in. So that should be the main motivation, because I believe what Habermas said, "Erkenntnis impliziert Interesse" [(i.e., insight implies interest)], and without being interested in the subject, we can forget about it. (I-2-15)

On the other hand, without high-control strategy components, the efficiency, quality, and coordination of work would be much more difficult to establish and maintain. High-control strategy components are needed to support the emergence of efficient, well-functioning, general systems and processes that can ensure that appropriate standards are developed and met by facilitating the ongoing reconciliation of individual and collective interests, as in the case of academic peer review. As one interviewee pointed out, academic peer review acts as a general quality assurance mechanism that sets the standards for appropriate work in a research community:

> What should be gone is that when people do something that has no scientific relevance or sound method and they just have an artifact, they stop saying: "This is DSR research!" and get through with it. Because that really drives me crazy when I get a paper as a reviewer that cites Hevner et al. (2004) and then they never talk about design and design aspects again. So that should stop. Because this is one thing that leads other people to say that DSR is not a valid method. (I-1-3)

This creates an indeterminate design situation with potential failure modes of over- and under-emphasis on control in complex real-world settings. For example, it is often argued that the academic system itself is prone to the emergence of perverse incentives, where measures and assessments of success drift away from what the academic community professes to value:

> It's a problem for science in general. And it goes back to Herbert Simon's combination of art and science, his basic idea in the 1960s. He published this very famous paper "The Business School Problem." [...] He said, [...] we're trying to [...] solve problems and advance knowledge. We're trying to be both scientifically and practically relevant. But in reality, we're not achieving this. And maybe it's unrealistic to achieve this in the first place. And often, as researchers, we end up doing what serves our community or what we like to do. [...] So we get together as academics at a conference and we present to each other. And because everyone is telling you how great what you're doing this year is, it reinforces you to continue doing it. It's like with the filter bubbles on Facebook. If I go to these places and I get this confirmation, it reinforces my confirmation bias. And I think I'm doing this relevant stuff. And because I am doing it, others are doing it. Well, the problem is that the intentions with which many people do this are not genuine knowledge development intentions at all times. Often they're just intentions [... for] getting a degree [or] getting money from a company. (I-2-12)

Against this backdrop, we frame the axiological challenge for responsible DSR as the design and implementation of robust and sustainable strategies that help to manage conflicting values and positions across the multiple levels of responsible DSR practice. This requires creating an appropriate mix of spaces for creative exploration, self-expression, and specialization around diverse issues (i.e., low-control strategy components), as well as more constraining coordination structures that exploit existing knowledge to focus collective resources on collectively valued goals and promote standardization around good practices (i.e., high-control strategy components). The end-in-view is to develop and maintain strategies and supporting infrastructures that facilitate the ongoing identification and renegotiation of shared values and positions, so that sufficient common ground can always be established for responsible DSR inquiry to be productive.

4.4 Methodological Challenge

Developing, Selecting, and Integrating a Portfolio of Practices

It has often been argued that DSR is a very complex, time-consuming, and unpredictable methodological approach to IS research (e.g., Vaishnavi & Kuechler, 2015). For example, in a keynote on how to do good DSR, Alan Hevner concluded his talk by stating that "daring to do good design science research requires the researcher and the research team to face these challenges of complexity, creativity, causality, confidence, control, contribution, cumulative knowledge, and collaboration" (Hevner, 2020). Similarly, interviewees in this study emphasized the complex and unpredictable nature of DSR: "Design processes are messy, you don't plan them perfectly ahead and you just follow the plan. Lots of things happen and things change" (I-1-9). This leads to the conclusion that "if you want to do decent [...] design science research, you should be aware of the fact that it takes time" (I-2-15). Statements such as these point to the methodological challenge of responsible DSR: the practical need to *develop, select, and integrate a portfolio of practices* that meet the contextual and often only vaguely defined but still evolving needs of the situation at hand.

In Figure 13, I frame this methodological challenge in terms of the need for *relevance realization* (Andersen et al., 2022; Vervaeke et al., 2012; Vervaeke & Ferraro, 2013a). Relevance realization is an emerging framework in cognitive science that seeks to explain how agents in open worlds are able to act intelligently, rationally, and wisely without succumbing to the frame problem—the recognition that combinatorial explosion imposes fundamental constraints on intelligent and rational action in an open world that changes through action over time (Andersen et al., 2022; Star, 1989; Vervaeke et al., 2012). For example, how can we rationally plan our actions when it is impossible for us to anticipate all the consequences of our actions? A key insight embedded in the relevance realization framework is the understanding that relevance is relational and therefore cannot be determined a priori, but must be constantly realized as agents change and are changed by their environment (Andersen et al., 2022). Thus, the relevance realization framework aligns well with our pragmatic understanding of responsible DSR, which emphasizes the open-ended and contextual nature of responsible DSR practice. It provides a coherent and parsimonious theoretical account of the key trade-offs that we face when acting in the world.

As Vervaeke et al. (2012) argue, relevance realization is best understood in the context of an open-world economic model, where resource scarcity leads to competing evolutionary imperatives for agents to exhibit behavioral *efficiency*, on the one hand, and behavioral *resilience* in the face of change, on the other (see also Andersen et al., 2022). While behavioral efficiency points to the need to focus investments on the most basic, general practices that ensure an agent's competitiveness and survival in the short term, behavioral resilience points to the opposite need to diversify investments across a variety of specialized practices that can ensure an agent's adaptability to changing conditions and promote long-term fitness. Both imperatives reinforce, but also constrain, each other in, what Vervaeke et al. (2012) call, a higher-order opponent processing relationship that can continuously evolve and adapt to the changing demands of the situation. In particular, the relevance realization framework emphasizes that agents in open worlds must constantly make three interrelated, action-oriented trade-off decisions regarding their behavioral portfolio:

- *Prioritizing investments* to balance risks and rewards (*diversifying vs. focusing*),
- *Scoping boundaries* to ensure applicability (*specializing vs. generalizing*), and
- *Tempering actions* to unlock long-term and short-term gains (*exploring vs. exploiting*).

When we apply this perspective to the development, selection, and integration of a portfolio of practices, we can see how responsible DSR requires us to *prioritize investments* to strike an appropriate balance between *diversifying* investments across a broad and resilient portfolio and *focusing* investments on the basic, general practices that help to maintain a productive flow of action, and thus competitiveness in the moment. We also need to decide how to *scope the boundaries* of the practices we invest in so that an appropriate level of *applicability* is maintained. This requires balancing the need for *specialized* practices (e.g., techniques), which have a limited scope but may be highly applicable in specific situations, with the need for *general* practices (e.g., integrative approaches), which are generally applicable in most situations but are less adapted to their specific needs. While behavioral resilience requires growing a large portfolio of specialized

Figure 13: Key Trade-offs for Managing a Portfolio of Practices (based on Vervaeke et al., 2012).

practices, resource constraints require maintaining behavioral efficiency, which calls for a limited portfolio focused on general practices. Effectively managing these trade-offs requires the *tempering of actions* over time, so that long-term and short-term interests are and remain appropriately balanced. In particular, *exploring* new practices can help build resilience and create opportunities that may pay off in the long term, while *exploiting* already mastered practices is a more efficient way to realize benefits in the short term.

The interviews confirm the centrality of these trade-off problems to the productive conduct of responsible DSR. For example, when considering the trade-off problem of prioritization (i.e., diversifying vs. focusing), the diversity of the DSR community was critically discussed, both in terms of its concerns and its practices. When asked for their definition of DSR, one interviewee said:

> That is a tough question. [...] It's a tough question [...] not because there's any lack of definitions, approaches, to DSR and so on, but [because] there is such diversity. (I-2-5)

While this diversity of DSR is well recognized (e.g., Rai, 2017a) and mostly viewed positively as a natural outcome of a pluralistic IS field that embraces novelty and change (e.g., Mendling et al., 2021), it may come at the cost of an inability to address complex, large-scale challenges that require focusing investments on fewer, but better-resourced projects and practices. As one interviewee pointed out, most DSR projects are very resource-constrained and therefore limited in what they can achieve:

> Most design science research is done by individuals or small teams, often involving novice researchers, like PhD students. There are obviously limits to what these people can do and what you can possibly achieve in that time frame. [...] So, a lot of the solutions that we've seen are necessarily small, prototype-ish, limited, and so forth, because at the end of it, they're [just] a bunch of guys that committed a few years to building stuff. (I-2-9)

At the same time, however, DSR can arguably be an excellent methodological framework for coordinating highly ambitious, resource-intensive, and large-scale projects:

> I use DSR to think of how to manage the structure of a very large project. By splitting it into smaller subprojects I can mix and match even more complex projects. So it is a very systematic approach for me to think. If

> I'm the conductor it gives me a way to coordinate the independent projects that are happening and everyone is doing their work quite independently but coordinated. [...] It can be used to tackle these kinds of really large problems. [But] I haven't seen that many papers on real-world applications. (I-1-2)

As another interviewee put it, DSR practitioners may be ideally positioned to help build bridges between diverse parties seeking to solve complex real-world problems:

> Very often, we need to team up to really build solutions. [...] Then what our quality could be, should be, probably, is really structuring the process. So, building bridges. Information systems would be very good at doing this in other areas as well. Facilitating dialog, creating boundary objects, modeling conceptualizations, structuring in order to orchestrate such a diverse endeavor. So, [I view] design science researchers and IS design science researchers really as those people tying all of those pieces together. (I-1-12)

This situation points to the untapped potential that can be cultivated through a more systematic engagement with what Gable (2020) calls "IS research strategy"—a deliberate discourse about the strategic priorities of the IS field that helps individual researchers and the IS community at large make more informed decisions about their investments of resources (e.g., money, time, energy, motivation, etc.). The goal is to explicit discuss trade-offs between priorities in order to find the right balance between focusing and diversifying our investments as a research community.

Facing the trade-off problem of scoping methodological boundaries for applicability (i.e., specializing vs. generalizing), the DSR community works, at one extreme, with general problem-solving practices that can be applied to almost any problematic situation. For example, DSR in the abstract is understood to follow a general problem-solving pattern that is similar across all problems and even across scales (i.e., the general problem-solving pattern applies not only to individual research projects but also to entire research streams; Vaishnavi & Kuechler, 2015). At the other extreme, we also work with and develop specialized techniques that encapsulate specific expertise on how to do a small task well (e.g., Avdiji et al., 2020). In practice, therefore, DSR requires traversing the full spectrum of applicability and integrating diverse practices into a coherent approach. At a conceptual level, this is achieved through a hierarchical nesting of DSR practices in terms of methods and techniques (see Wieringa, 2009):

> A technique is a [specialized] problem solving mechanism that is reused in several methods. A method is a systematic way of [combining techniques to achieve ...] certain results. (I-1-7)

Due to the flexibility of hierarchical nesting, DSR is characterized by a highly eclectic and context-sensitive use of methods and techniques:

> It will involve all sorts of techniques and methods. For instance, you might need quantitative data analysis at some point. You might need a statistics package for instance. You might do interviews, let's say in a focus group, or when you evaluate an artifact, or when you do requirements engineering in order to come up with a conceptualization in the first place. (I-1-10)

This has created a challenging situation for the DSR community, because as more forms of DSR emerge from different strategies of combining methods and techniques (e.g., Iivari, 2015; Rai, 2017a), it has become difficult to coordinate around appropriate schemas for evaluating DSR practice (e.g., Peffers et al., 2018). When there are arrangements of methods and techniques that differ significantly in character, it becomes important, but often difficult, to find reviewers who are accustomed to approaching a problem in a similar way. As one interviewee put it:

> Because of this diversity in DSR, it is sometimes very difficult for people outside to understand what it is, because if you look at the different flavors, [...] it's just like what holds this stuff together? [...] So, I think that this can be problematic when somebody who is writing an article based on DSR, trying to publish it, does not align with a reviewer or a review team that is reviewing that article and has a different conception of what DSR is. That's a problem. (I-2-5)

The extent of this problem is illustrated by another interviewee, who explains that some researchers whose work could clearly be described as DSR do not use the term because we have not yet developed a common understanding of how the different forms of DSR relate to each other, which calls into question the legitimacy of the DSR paradigm as a whole:

> So, [this researcher] does design science research. The thing is that he doesn't call it design science research. It is a specific type of design science research. [...] But he never says in his papers that he is doing design science research and he doesn't build upon Hevner or Gregor or Herbert Simon or any of those kind of people. Right. [...] There's no shared understanding. It's not like with qualitative research. If we say case study, most people will know more or less what is a case study and people will say there is Robert Yin and there is this and there is that. So, there is like a cumulated, shared understanding in the community about [...] what case study research is. [...] Now, with design science research it's very diverse. You have people who are doing design science research that is of a certain type that has nothing to do with the type of design science that other people are claiming to do. So it's like there's no consensus and that doesn't help. Because if there's no consensus and shared understanding about design science research, then it doesn't help to establish the legitimacy of this paradigm in the IS community. (I-2-12)

However, even agreement on methods and practices would not be sufficient to ensure proper evaluation of DSR efforts. As one interviewee pointed out, in-depth evaluations also require considerable experience with the problematic situations under study:

> One of the problems in the IS field is that a lot of people don't have a way of evaluating what they see, because a lot of people really don't have a lot of background in topics that they're reviewing on. So, these checklists end up being a fairly useful thing for them, but a checklist as something... Can you imagine, you went to a doctor, and the doctor actually, really wasn't trained as a doctor, but he had a checklist? You know, that might be okay if the question was: "you're drinking too much, don't drink so much." But if the question is: "you have some kind of unusual problem." That's going to be ridiculous. You know, it's not going to get you anywhere. (I-2-13)

This challenge of relevant expertise runs deep and extends to the field of DSR education. Because DSR tends to require exposure to a very broad range of topics, methods, and techniques, it becomes difficult to decide what content to teach and to what extent:

> What do we teach? Because it's so diverse, right? It's not simply one area. It can be different areas. And each area has got its own traditions and so on. So, what do we teach? (I-2-7)

All of these observations point to the challenges associated with the scoping of methodological boundaries in DSR practice. Because DSR is essentially a general problem-solving practice that is focused on integrating specialized methods and techniques to address wicked problems, it must be inherently more diverse than research approaches with less open-ended goals (Ashby, 1968). Reaching agreement on how to productively manage this methodological diversity will be one of the major challenges for a responsible DSR ecosystem in the future.

Reflecting the trade-off problem of tempering actions for both short-term and long-term gains (i.e., exploring vs. exploiting), the interviews highlight the role of uncertainty in the temporal dimensions of DSR practice. For example, some interviewees suggest that DSR tends to be more involved and longer lasting than more traditional forms of research:

> And maybe a difference between classic behavioral research and design science research is that a design science research project lasts longer. There is more knowledge that is somehow interrelated. So, if you are investigating the effect of [x] on [y], you do one field study or an experiment and this is then one closed part [. . .] and if you then do another study [. . . that is a different study]. In design science all these somehow build upon each other. (I-1-3)

As a result, many researchers view engagement in DSR as inherently riskier than more traditional forms of research, and thus potentially less attractive in the short term (see also Tremblay et al., 2018):

> A lot of people told me: "Why do you do DSR? [. . .]" Doing DSR, you always have more steps to make. And this is hard, and this hard to put into a paper, and this is more of a long term thing. So this is why it's hard to find. (I-1-8)

How to make DSR more accessible, less costly, and easier to pick up and publish has therefore been a key concern of many DSR publications to date (e.g., Baskerville et al., 2018; Gregor & Hevner, 2013; Hevner et al., 2004; Peffers et al., 2018; Peffers et al., 2007; Sein et al., 2011; Vaishnavi & Kuechler, 2015). However, as one interviewee pointed out, if DSR is to address real-world problems, a long-term perspective may be unavoidable to allow research efforts to reach the scale necessary to have a positive impact:

> I think it's very important if you're doing research and you would like to develop a research line that you should be in it, let's say, not for the short-term perspective but for the long-term perspective. [. . .] I see some professors that whenever industry says something, that is what they start doing. It is very difficult to build out a research line on that. I think we need to be consistent, really think about this long-term. I think that, in the end, if you will do things at a larger scale, you need to have funding, right?[. . .] If you want to do design science where you're trying to tackle real-world problems, you typically need to have some scale, and that scale you only get through funding in some way. (I-2-25)

Overall, finding effective ways to link the exploration and exploitation of appropriate practices over time is a complex challenge that benefits from being viewed holistically, not only from an individual or project perspective, but also from the perspective of the overarching DSR ecosystem (e.g., Herwix & Rosenkranz, 2019; Morana et al., 2018). The relevance realization framework, which I used to frame the methodological challenge, provides a promising starting point for such a holistic engagement. The competing evolutionary imperatives of behavioral efficiency, on the one hand, and behavioral resilience in the face of change, on the other, and the trade-offs that must be made because of them, point us to considerations that can help us be more deliberate and responsible in the development, selection, and integration of our portfolio of DSR practices.

5 Recommendations for Developing a Responsible Design Science Research Ecosystem

I have developed four recommendations (summarized in Table 8) that are designed to address the four paradigmatic challenges of responsible DSR in an integrated way and to promote the development of a responsible DSR ecosystem—one that balances the need for innovation with the need to ensure that these innovations are beneficial to society as a whole. They are based on a deep engagement with the interviews, the prior literature, and the framework for responsible DSR. While I do not intend to position these recommendations as a panacea that will help us anticipate and address all the challenges we may encounter on our path to a sustainable, responsible DSR ecosystem, I do argue that they provide a useful and well-informed starting point for realizing this vision.

Table 8: Recommendations for Developing a Responsible DSR Ecosystem.

Recommendation	**Description**
Establish Systematic Boundary Discourse	Systematic boundary discourse refers to the systematic use of boundary critique to guide us and our stakeholders in reflecting on and consciously (re)constructing our boundary judgments in responsible DSR practice. It can help us to become more responsible in the identification and selection of our research problems, as well as in the development and evaluation of our research projects.
Support Responsible Multi-Layered Development	Responsible multi-layered development is a high-level DSR process that views responsible DSR as a collective activity involving an entire research ecosystem. It identifies four layers of research that build on top of each other as responsible DSR moves towards resolving problematic situations. It aims for a more responsible coordination between relevant parties and a more productive allocation of resources.

Recommendation	Description
Facilitate Polycentric, Multi-level Cooperation and Value Co-creation	How to effectively facilitate polycentric, multi-level cooperation and value co-creation at scale is not only one of the most pressing societal issues of our time, but also a question we must find an answer to if we are to develop a sustainable, responsible DSR ecosystem. I suggest that learning-by-doing in the form of the development and ongoing maintenance of a responsible DSR ecosystem is an appropriate way forward. Several existing insights and frameworks can be drawn upon to support this endeavor.
Develop Pragmatism, Reflexivity, Sensitivity, and Anticipation	Responsible DSR is a program that requires a strong methodological portfolio and the necessary skills and understanding to use it productively. Pragmatism provides a proven theoretical foundation and motivation for these efforts. Reflexivity opens up new perspectives that support its progress. Methodological and theoretical sensitivity are key competencies it requires. Anticipatory thinking and responsive action are what it seeks to develop.

5.1 Recommendation 1

Establish Systematic Boundary Discourse

The first recommendation is to work towards establishing *systematic boundary discourse* as a core component of a responsible DSR ecosystem. This recommendation is inspired by Ulrich's (2003; 2001, 2006) work on systematic boundary discourse, which suggests a critical pragmatic solution to the challenge posed by the inherent selectivity of our framings. If even the most conscientious and responsible professionals cannot fully justify their decisions and conclusions, because all our framings are always incomplete and provisional, then we should continue to engage in critical discourse until, ideally, a state is reached that is acceptable to all concerned (see the section on ontological challenge).

In particular, the core idea behind systematic boundary discourse is that we should systematically engage in *boundary critique*: the practice of making explicit the selectivity of our framings and opening them up to critical reflection and discourse (e.g., Córdoba & Midgley, 2008; Ulrich, 2003; Ulrich, 2001; Ulrich & Reynolds, 2010). Several frameworks, tools, and practices for facilitating boundary critique have been developed and found useful for gaining a systematic understanding of our boundary judgments and for building collective agreement on how to frame problematic situations and potential solutions (e.g., Córdoba & Midgley, 2008; Sydelko et al., 2021; Ulrich, 2003; Ulrich, 2001; Ulrich & Reynolds, 2010). While boundary critique is not a silver bullet for resolving the challenge posed by the inherent selectivity of our framings, it does offer a pragmatic, systematic way to address it through critical discourse (Ulrich, 2003; Ulrich, 2001, 2006).

Table 9 lists a set of boundary questions developed by Ulrich (e.g., 2001, p. 96) that help to illustrate the potential value of systematic boundary discourse for a responsible DSR ecosystem. The boundary questions support a holistic, yet parsimonious, systematic reflection on the *sources of motivation*, *power*, *knowledge*, and *legitimation* that condition our framing of projects, programs, or other organizational systems. As such, they provide a general checklist that we can use to systematically develop, evaluate, and summarize responsible DSR efforts along key dimensions that integrate and reflect the challenges identified by our paradigmatic framework for responsible DSR.

Table 9: Boundary Questions of Critical Systems Heuristics (e.g., Ulrich, 2001, p. 96)

Dimension	Boundary Question
Sources of Motivation	Who is (ought to be) the **client**? That is, whose interests are (should be) served?
	What is (ought to be) the **purpose**? That is, what are (should be) the consequences?
	What is (ought to be) the **measure of improvement**? That is, how can (should) we determine whether and in what way the consequences, taken together, constitute an improvement?
Sources of Power	Who is (ought to be) the **decision maker**? That is, who is (should be) in a position to change the measure of improvement?
	What **resources** are (ought to be) controlled by the decision maker? That is, what conditions of success can (should) those involved control?
	What conditions are (ought to be) part of the **environment**? That is, what conditions does (should) the decision maker not control (e.g., from the viewpoint of those not involved)?
Sources of Knowledge	Who is (ought to be) involved as a **professional**? That is, who is (should be) involved as an expert, e.g., as a system designer, researcher, or consultant?
	What **expertise** is (ought to be) consulted? That is, what counts (should count) as relevant knowledge?
	What or who is (ought to be) assumed to be the **guarantor**? That is, what is (should) be considered a source of guarantee (e.g., consensus among experts, stakeholder involvement, support of decision-makers, etc.)?
Sources of Legitimation	Who is (ought to be) **witness** to the interests of those affected but not involved? That is, who is (should be) treated as legitimate stakeholder, and who argues (should argue) the case of those stakeholders who cannot speak for themselves, including the handicapped, the unborn, and non-human nature?
	What secures (ought to secure) the **emancipation** of those affected from the premises and promises of those involved? That is, where does (should) legitimacy lie?
	What **worldview** is (ought to be) determining? That is, what different visions of improvement are (should be) considered and somehow reconciled?

In particular, the boundary questions can be used in two modes—is and ought—and combined with the three key dimensions of problem and solution spaces (i.e., level of abstraction, logical scope, time horizon) that I have identified as part of the framing of the ontological challenge to systematically plan, reflect, critique, and position the motivation of responsible DSR efforts. For example, the mobile health app development project mentioned earlier (see the section on the ontological challenge) illustrates how a critical reflection on the logical scope of the problem space can lead to a shift in the measure of improvement and, subsequently, the researchers' view of the client and even the purpose of a project. While neither the boundary questions nor the paradigmatic framework were needed to recognize the value of a framing shift in this example, their combined use can prepare researchers to arrive at such insights more reliably.

As several examples illustrate, boundary critique exercises are a highly useful means of exploring, reflecting on, and aligning framings in multi-actor settings (e.g., Córdoba & Midgley, 2008; Ulrich & Reynolds, 2010). For example, Córdoba and Midgley (2008) describe how such exercises have supported the development of a more comprehensive and inclusive problem framing in a large-scale IS development project. Here, boundary critique exercises helped to more productively include the voices of parties that would otherwise have been marginalized in the project planning. More generally, Alter's (2013) work on Work Systems Theory, and in particular his recommendation to use work systems snapshots to align relevant parties in IS development projects, and the rapid adoption of visual inquiry tools (e.g., the business model canvas; Avdiji et al., 2020; Osterwalder & Pigneur, 2010) for the collective exploration of design options and opportunities, provide additional evidence of the potential value of multi-actor boundary critique exercises when exploring, reflecting, and aligning framings in creative design processes.

Despite this potential value, there is a major academic challenge associated with the systematic use of boundary critique: how to faithfully communicate the resulting shifts in framing in research papers and reports. As the interviewee involved in the mobile health app development project reflected:

> I think that's the hard part in design science. You're solving real problems, and when you go see these problems in the field, the measure of success changes, and does it fit nicely within one academic paper? I'm not sure. It's really hard to tell that story. (I-2-2)

Systematic boundary discourse, informed by the boundary questions in combination with the paradigmatic framework presented in this paper, provides a productive way to address such communication issues. The boundary questions can be used to create holistic, yet parsimonious, snapshots of responsible DSR efforts at specific points in time, which in turn can document the evolution of such efforts in a systematic and efficient manner. It would be easy and inexpensive to attach one or more of such snapshots as appendices to publications (see the Online Appendix for an example based on this paper). This would seem particularly helpful for maintaining more systematic documentation and allowing for a more comprehensive reflection of long-term DSR projects, programs, or streams that go through multiple iterations over several years (e.g., Paradice et al., 2019). It would also respond to calls for more explicit reporting of the values that guide our research efforts, which have recently been raised in an effort to promote greater ethical awareness in our practice (Herwix et al., 2022).

Given the potential benefits to be gained and the low costs and risks associated with establishing a more systematic boundary discourse, I argue for the creation of appropriate incentives to encourage the systematic adoption of boundary critique exercises in responsible DSR practice. This could take the form of educating peer reviewers to request explicit documentation of key boundary judgments underpinning the evaluated work, journals advocating such documentation as part of their transparency policies (e.g., Burton-Jones et al., 2021), or even journals offering tailored review processes for registered reports (Chambers & Tzavella, 2020) that are informed by boundary critique exercises. Moreover, all interested parties are encouraged to take advantage of the emancipatory opportunity offered by the use of boundary critique exercises. Since all our framings are always incomplete and provisional, we always have a sound logical and moral justification for challenging potentially problematic boundary judgments made by others around us (Ulrich, 2006). The boundary questions provide a simple checklist that can be used to begin asking the right questions. With this in mind, we should also accept our own responsibility to remain open to critically examining and revising our boundary judgments, lest we become dogmatic.

5.2 Recommendation 2

Support Responsible Multi-Layered Development

The second recommendation is to move toward a *multi-layered development model for responsible DSR* (MLDM) that explicitly recognizes the collaborative and discursive nature of research practice. While existing DSR frameworks tend to focus on the development of individual research projects (e.g., Peffers et al., 2018) or position DSR in the context of IS research at large (e.g., Hevner et al., 2004; Seidel & Watson, 2020), the MLDM recognizes responsible DSR as a collaborative problem-solving effort that depends on and involves an entire research ecosystem. Such a shift in perspective is necessary if we are to enable an effective and responsible coordination among relevant parties and a productive resource allocation within a responsible DSR ecosystem.

The core idea of the MLDM (see Figure 14) is that responsible DSR is a discursive practice (e.g., Ulrich, 2003; Ulrich, 2001) that emerges from deliberate, open-ended engagement with problematic situations and follows a common problem-solving pattern (e.g., Dewey, 1938b; Farrell & Hooker, 2013; Vaishnavi & Kuechler, 2015). This understanding of responsible DSR is *scale-invariant* in the sense that it can be applied not only to individual research projects, but also to large-scale inter-, multi-, or transdisciplinary research programs (e.g., Nunamaker et al., 2017), and even to research streams that integrate the results of many distributed research efforts (e.g., W. Kuechler et al., 2005; Paradice et al., 2019). In principle, this allows the model to act as a boundary object that can help bridge between these previously mostly isolated perspectives and facilitate a general understanding of what it means to do responsible DSR. Specifically, the MLDM distinguishes between four *layers* and four *stages* of responsible DSR.

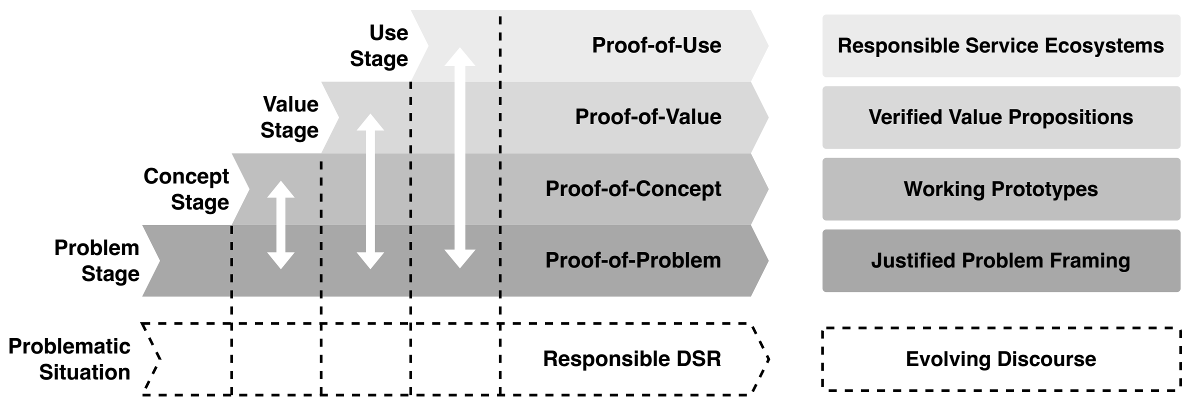

Figure 14: Multi-Layered Development Model for Responsible DSR.

The four *layers of responsible DSR* indicate a broadening of the focus of inquiry as problematic situations move closer to resolution. While any responsible DSR effort should begin with a focus on investigating and appropriately framing the problematic situation that triggered it (i.e., *proof-of-problem*), over time additional considerations are layered on top: Prototyping potential solutions (i.e., *proof-of-concept*), evaluating value propositions (i.e., *proof-of-value*), and responsibly deploying effective solutions (i.e., *proof-of-use*). These layers are hierarchical in the sense that higher layers build on the insights generated by lower layers. For example, prototyping potential solutions requires a problem framing that is already justified to some degree. However, the lower layers are also informed by work on the higher layers. By prototyping potential solutions, we gain important insights that can help us reframe the problem situation. Therefore, the different layers are not isolated from each other, but should be brought into dialog whenever possible.

Given this understanding, maintaining an open and critical discourse about ongoing developments at each layer becomes the key to facilitating effective and responsible coordination within a responsible DSR ecosystem. I suggest that this could best be realized through an open, community-wide discourse at each of the layers, driven by multiple layer-specific publications per project (i.e., a step-by-step model). This stands in contrast to the current practice in IS research, which requires individual studies to attempt to address individual problems across as many layers as possible (i.e., a one-study-does-it-all model).[14] Thus, instead of setting the bar for publication based on the extent of layer coverage, I propose that we incentivize and enable a greater focus on the quality of our work at each layer through earlier publication and ongoing feedback cycles. This discourse model has the advantage of leading to less publication bias, more systematic reporting of work done and insights gained, and more responsible co-opetition in the development of useful knowledge.

The four *stages of responsible DSR* support this discourse model by acting as checks and balances to ensure responsible progress towards resolution. As emphasized in the section on the epistemological challenge, responsible DSR must facilitate a productive interplay between knowledge-as-object and knowledge-as-process in order to increase the chances of virtuous cycles and minimize the risks of vicious cycles for all concerned. Previous research argues that one potential strategy to achieve this is to pursue *differential intellectual progress*, which is to "prioritiz[e] risk-reducing intellectual progress over risk-increasing intellectual progress" (Muehlhauser & Salamon, 2012, p. 17). While the appeal of differential intellectual progress may be clear to responsibility-minded actors, securing it is a difficult undertaking. Because much intellectual progress is inherently dual-use (i.e., it can be used for risk-reducing as well as risk-increasing purposes) and may come at a higher cost if it is pursued more responsibly (e.g., Askell et al., 2019), appropriate incentives and mechanisms must be put in place to ensure differential intellectual progress (e.g., Consilience Papers, 2022). I argue that a rigorously enforced, stage-gate-based approach to development could be at least part of the solution. In particular, the MLDM proposes four stages that responsible DSR should pass through as it progresses toward resolution (cf. Nunamaker et al., 2015; Vaishnavi & Kuechler, 2015):

1. The *Problem Stage* can be passed with a justified proof-of-problem,
2. The *Concept Stage* can be passed with a proof-of-problem and a working proof-of-concept,
3. The *Value Stage* can be passed with a proof-of-problem, a proof-of-concept, and a verified proof-of-value, and
4. The *Use Stage* is open-ended, with the goal of monitoring and maintaining a proof-of-problem, proof-of-concept, and proof-of-value at all times, while responsibly deploying

[14] For example, "the current structure [...] provides scant opportunity to publish work that seeks to substantiate particular problems" (Winter & Butler, 2011, p. 106), which is a problem because "coming up with the right answer to the wrong question does not create value" (Rai, 2017b, p. iii).

a solution to achieve a proof-of-use.

I acknowledge that enforcing these four stages, for example by requiring the submission of registered reports for the necessary proofs at each stage for peer review, would place stringent demands on responsible DSR practitioners. However, given the potential for unintended catastrophic downsides from the unrestrained development of digital technologies (e.g., Consilience Papers, 2022; Harari et al., 2023; Muehlhauser & Salamon, 2012; Yudkowsky, 2023; Zuboff, 2015), the cost seems a worthwhile investment. For example, such a model could help to address the *Unilateralist's curse* (Bostrom et al., 2016), which emphasizes that errors in individual judgments can lead to risky projects being undertaken more often than would be desirable—even when all parties involved are trying to act in everyone's best interest. As Bostrom et al. (2016) themselves point out, deferring decisions on whether to pursue risky projects to expert consensus (as could be implemented through peer review) can be a useful mechanism for lifting the curse.

Furthermore, iterative and ongoing critical reflection and peer review, as envisioned in the MLDM, could also help address Collingridge's *dilemma of control* (e.g., Collingridge, 1980; Genus & Stirling, 2018). While the dilemma points to a fundamental challenge for responsible innovation (Genus & Stirling, 2018), de Reuver et al. (2020) argue that the dilemma can be overcome if critical evaluations and reviews are integrated throughout the technology development lifecycle and sufficient flexibility for critical revision is built into the development effort from the outset. The MLDM is designed to do just that.

In addition to strengthening our ability to take responsibility for our research efforts, implementing the MLDM can also help increase our collective productivity by leading us to establish rigorous feedback cycles earlier in the research process. If we invest more of our collective resources in understanding and appropriately framing the problematic situations we face, rather than jumping to prototyping solutions, we could begin to more systematically identify the most promising opportunities for responsible DSR to make a positive difference (see Gable, 2020; Herwix et al., 2022). This would allow us to focus our resources more deliberately, develop more collective expertise on important issues, and avoid unnecessary "Type III" errors (Rai, 2017b)—investing resources in research problems that do not matter.

Overall, the MLDM argues for a significant departure from conventional wisdom on how to structure DSR efforts. However, given the rapidly accelerating pace of technological development and the associated risks, the time has come for a critical re-examination of our established modus operandi. The MLDM can inspire and usefully inform such an effort.

5.3 Recommendation 3

Facilitate Polycentric, Multi-level Cooperation and Value Co-creation

The third recommendation is to create institutional arrangements that facilitate *polycentric, multi-level cooperation and value co-creation* within our research ecosystem, using the best frameworks, methods, and tools available or even developing our own. It is based on the five axioms of service-dominant logic defined by Vargo and Lusch (2016, p. 18), which state that "service is the fundamental basis of exchange" (Axiom 1), as "all social and economic actors are resource integrators" (Axiom 3) who "uniquely and phenomenologically" determine the value of the service (Axiom 4) they "cocreate" together with multiple actors (Axiom 2), using "actor-generated institutions and institutional arrangements" for coordination (Axiom 5). In this perspective, facilitating cooperation and coordination among relevant parties becomes a core concern, as it is recognized as a necessary condition for mutual service exchange and value co-creation to occur. This is achieved through the development of institutional arrangements that link resource-integrating actors in relatively self-contained, self-regulating systems of mutual service exchange, so-called *service ecosystems* (Lusch & Vargo, 2014a; Vargo & Lusch, 2016).

Service ecosystems are key to value co-creation because they institutionalize mechanisms and constraints that reduce the amount of cognitive resources required to engage in mutual service exchange, and also enable network effects, where more actors sharing a service ecosystem means greater potential coordination benefits for all (Vargo & Lusch, 2016). To work well, however, service ecosystems need to be organized around shared purposes (e.g., Schlaile et al., 2023; D. S. Wilson et al., 2023; D. S. Wilson et al., 2013). This means that, in practice, service ecosystems tend to be nested and overlapping, with multiple formally independent centers of decision making that may or may not take each other into account (Vargo & Lusch, 2016)—a situation that Ostrom (2010, p. 643) refers to as *polycentric*.

In order to effectively implement a responsible DSR ecosystem, I suggest drawing on the work of D. S. Wilson et al. (2013), who generalized Ostrom's (2010, 2015) work on the governance of common pool resources to derive a set of eight core design principles (CDPs) that empirically characterize well-cooperating and sustainable groups (i.e., service ecosystems) (see Figure 15). The CDPs provide a versatile framework because, by virtue of their alignment with multilevel cultural evolution theory, they can arguably inform the design and development of polycentric service ecosystems at all scales, from small groups to large organizations, cities, nations, and ultimately the entire Earth system (D. S. Wilson et al., 2023). In particular, they can serve as a useful checklist of key concerns that should be considered in the design and development of polycentric service ecosystems. For example, CDP 1 highlights that a sustainable service ecosystem requires a common purpose that motivates the participating actors to engage in mutual service exchange. By recognizing this requirement, actors in service ecosystems can begin to consciously reflect on the purposes that drive their interactions and more deliberately develop supportive institutional arrangements. To this end, CDPs 2-6 provide more specific guidelines, each of which contributes to the creation of a more cooperative environment in which disruptive, self-serving behaviors are suppressed in favor of purposeful, mutual service exchange and value co-creation. The CDPs 7-8 make the framework multi-level by extending the same principles to interactions between service ecosystems.

In addition to their use as a simple checklist, the CDPs have also been translated into measurement tools that can capture the extent to which the CDPs are being implemented by relevant parties in a particular context (D. S. Wilson et al., 2020). When used in this way, the CDPs can provide a holistic snapshot of the current "cooperative climate" in a particular service ecosystem, which can be used not only to document the evolution of the service ecosystem over time, but also to inform its development in the future.

For the management and development of sustainable service ecosystems, methodological approaches have emerged that combine the CDPs with evidence-based behavior change interventions to provide more systematic guidance and support (e.g., Schlaile et al., 2023). The core idea of such approaches is to enable complex, polycentric service ecosystems, in which it is easy for actors to act at cross-purposes, to come together in a way that makes them adaptive *as a system* (D. S. Wilson et al., 2023). In this view, the key to cooperation and value co-creation is to organize multi-level processes of "system-level cultural evolution in which all three ingredients—the target of selection, variation oriented around the target, and replication of best outcomes—are carefully managed [with the whole system in mind]" (D. S. Wilson et al., 2023, p. 4).

Given this understanding, I argue that the CDPs are particularly well suited to inform the development of a responsible DSR ecosystem, which must bring together actors and

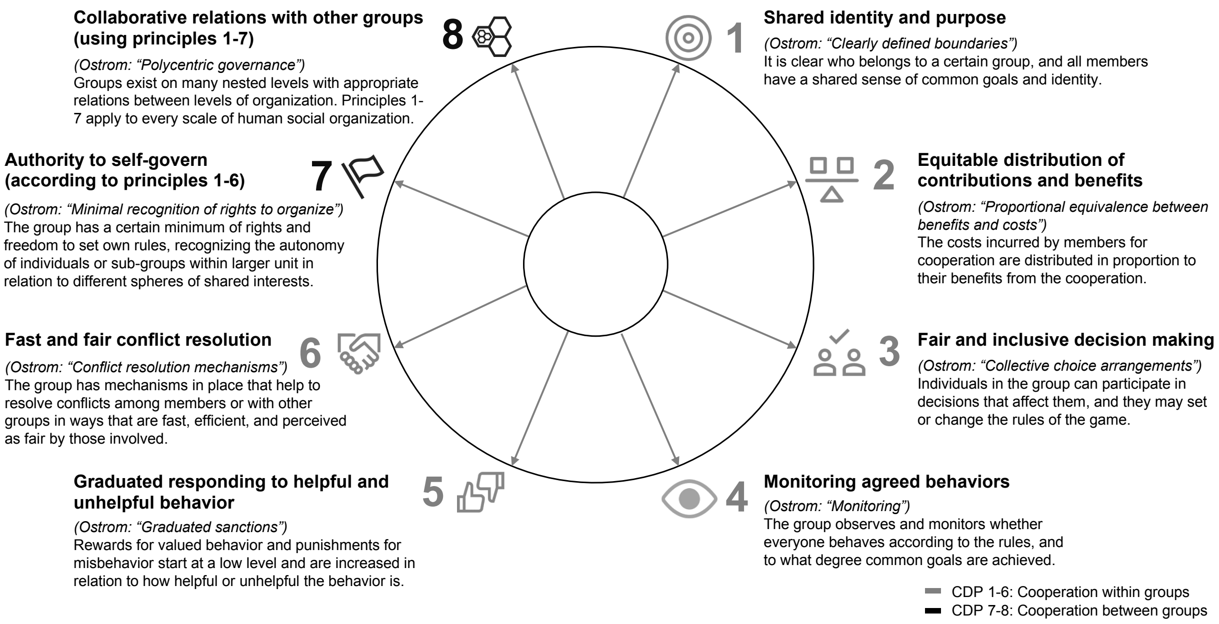

Figure 15: The Eight Core Design Principles (based on Schlaile et al., 2023).

initiatives with a variety of (potentially competing) interests and at a wide range of scales, from small research projects to large inter-, multi-, or transdisciplinary research programs, and even distributed research streams. As discussed in the section on the axiological challenge, a key issue in this context is how to develop appropriate strategies for managing conflicting values and positions, since conflicts are bound to arise sooner or later. In particular, we need to find management mechanisms that implement control in a way that allows the responsible DSR ecosystem to become adaptive *as a system* that is both efficient and resilient at the same time. While the CDPs do not provide a panacea for all the problems that will arise in addressing this task, they do provide a useful framework that can help us develop contextually appropriate responses. As D. S. Wilson et al. (2023, p. 4) argue, there can be no master plan for sustainable cooperation and value co-creation that is applicable in every situation; it is a matter of organizing context-appropriate processes of system-level cultural evolution.

Consider, for example, the publishing system as the heart of the IS research ecosystem. It is constantly evolving, as evidenced by the recent addition of three new journals to the Association for Information Systems' (Association for Information Systems, 2023) "Senior Scholars' List of Premier Journals" (SSLoPJ),[15] or the recent launch of a transparency initiative at MIS Quarterly aimed at supporting open science practices (Burton-Jones et al., 2021). However, to what extent is this evolution being carefully managed with the whole system in mind, or is it "simply" an emergent product of competitive interactions in a complex system? While it is beyond the scope of this paper to explore this question in depth, I will briefly illustrate how the CDPs could help to support the former. For example, they could be used to assess the "cooperative climate" at the level of the system as a whole, and also at the level of particular journals or conferences, based on a survey of a wide range of relevant parties, such as editors, reviewers, authors, practitioners, or even grant makers. This would help to explore key questions such as the extent to which the publication system actually has a common purpose (CDP 1), the perceived fairness of the distribution of costs and benefits (CDP 2), and the perceived inclusiveness of decision-making practices

[15] As its name suggests, this list aims to identify the top journals in the IS field. By doing so it intends to provide orientation to researchers interested in the IS field and "more consistency and meaningfulness to tenure and promotion cases" (Association for Information Systems, 2023).

(CDP 3). Once we have answers to these questions, it would be possible to develop context-appropriate action plans to flesh out how the stakeholders in the system can come together to seize the opportunities identified and address the remaining shortcomings. The key point of such an exercise would be to broaden the scope of our thinking to the whole system and help us adopt a flexible, cooperative mindset in assessing the current situation and planning our next steps. For example, ambitious changes to the publication system, such as inverting the publication model from pre-publication to post-publication review (Hyde et al., 2022), are likely to require coordinated efforts by multiple parties (e.g., journals or conferences) simultaneously, as unilateral action may be perceived as too risky (e.g., Berenbaum, 2023). In addition, the CDPs can also be used within such large-scale, system-wide efforts to align smaller (sub)groups of relevant parties, such as research teams or editorial boards of journals and conferences (cf. Schlaile et al., 2023). As noted above, the CDPs are not a panacea, but they do provide a useful roadmap that can help us chart our path to a more sustainable and cooperative future.

A final aspect to keep in mind in this context is the dual role that IS researchers play in the development of their own research ecosystem (cf. D. S. Wilson et al., 2023, p. 7). On the one hand, we act as responsible designers who must have the welfare of the whole system in mind. On the other hand, we are also self-interested participants in the systems that we design. This situation can create tension, because while as participants we have the right to respond to our own needs, as responsible designers we must recognize that our values and positions are only one among many. Therefore, we should ensure that whatever we put into practice is critically examined from multiple perspectives and is a benefit rather than a risk to the common good (see recommendations 1 and 2).

5.4 Recommendation 4

Develop Pragmatism, Reflexivity, Sensitivity, and Anticipation

The fourth recommendation is to strengthen the skills and methodological toolbox for responsible DSR by developing our capacities for *pragmatism*, *reflexivity*, *sensitivity*, and *anticipation* (see Table 10). These development opportunities address each of the four *ways of knowing* that were introduced in the section on the epistemological challenge. In doing so, we seek to outline ways in which responsible DSR practitioners can grow in a holistic and integrated manner. Similar to an interviewee, I argue that education and professional development play a crucial role in a successful responsible DSR ecosystem:

> If design science research helps [...] to train and develop new information systems professors who will later do important jobs in terms of educating people out there, then design science research can be as impactful as it can. But that will only happen if we do it systematically and people are trained to think rigorously. Because if people don't think rigorously in the process of doing design science research, they don't develop themselves and then they don't become great professionals, who then become better at teaching and having a great impact or doing better advising on policy. (I-2-12)

Table 10: Development Opportunities for Responsible DSR.

Opportunity	Ways of Knowing	Description
Pragmatism	Participatory	Pragmatism has been positioned as a worldview that is rooted in "the ongoing act in context" (Hayes et al., 1988, p. 100). It is flexible in the sense that it can work with other worldviews. Thus, pragmatism is an excellent worldview for open-minded researchers who want to take advantage of the strengths and overcome the weaknesses of different worldviews in order to approach problematic situations from multiple perspectives.
Reflexivity	Perspectival	Reflexivity is concerned with how we relate to the world and what that produces for ourselves and others (e.g., Serra Undurraga, 2021). It is synergistic with open-minded and open-ended participatory inquiry. It can be nurtured and developed at individual and collective levels.
Sensitivity	Procedural	Sensitivity refers to a researcher's ability to detect relevant signals in their environment and develop contextually appropriate responses (Bryant, 2017; Glaser & Strauss, 1967). It is critical in dynamic, complex, and changing situations, as is common in responsible DSR.
Anticipation	Propositional	Anticipation is the ability to imagine "how unexpected events may affect plans and practices" (Klein et al., 2010, p. 235). It is associated with a mode of thinking that is particularly relevant to responsible DSR, as it can enable effective action even in highly uncertain situations with potentially catastrophic outcomes.

Beginning with a focus on *participatory knowing*, I suggest that *pragmatism* offers an appropriate worldview for people who are interested in putting responsible DSR into practice. The key difference between pragmatism and other worldviews is its contextual nature, rooted in a focus on "the ongoing act in context" (Hayes et al., 1988, p. 100). In the light of pragmatism, nothing is inherently true or necessarily permanent, since the meaning of concepts is determined in action and application (Dewey, 2007). Ultimately, judgments should have *warranted assertibility*, that is, they should be the result of successful inquiry into the situation at hand (Dewey, 1941). As such, pragmatism overcomes the problems of naïve realism (i.e., only one way of representing the world is true) and caricature relativism (i.e., all ways of representing the world are true) by recognizing that truth is determined by what works in inquiry, and thus no predetermined positions are warranted (Dewey, 1938b; Hayes et al., 1988). In this inherently dynamic worldview, change is always possible, as new insights may require a reevaluation of previous judgments (Dewey, 1938b, 2007).

Pragmatism is well suited to ground and guide responsible DSR practitioners because it offers an open-minded and open-ended perspective that seeks to work with other worldviews and does not prejudge competing concerns raised in problematic situations. In particular, rather than superimposing one's own preferences and desires on problematic situations, it mandates a commitment to inquiry in order to find solutions that work for all concerned. Since this is not always possible (e.g., because of conflicting values and positions; see the section on the axiological challenge), it generally requires confronting the limits of inquiry itself as a formative part of inquiry (Dewey, 1938b; Ulrich, 2006). In a sense, pragmatism can do no more, but also no less, than advocate for placing the Golden Rule (Flew, 1979) at the center of a worldview—"do unto others as you would have them do unto you". This also makes pragmatism a highly motivating worldview because it implies room for open-ended exploration, experimentation, and the gaining of new insights, since all experience is recognized as inherently interactive and unique (Dewey, 1938a, 1938b).

A key feature of pragmatism is that it can work together with other worldviews insofar as their perspectives help to advance inquiry toward the resolution of problematic situations (Hayes et al., 1988). Thus, it seeks to transcend paradigmatic debates by recognizing that all worldviews have strengths and weaknesses and by advocating that they be used in a context-appropriate manner. To rephrase a common aphorism: *All worldviews are limiting, but some are useful some of the time.* This open-minded perspective allows pragmatists to leverage the strengths and overcome the weaknesses of different worldviews as they work to resolve problematic situations. The key is not to stick to one's guns, but to use what is available to make progress toward a successful resolution. In this sense, pragmatism is the ideal worldview for people in leadership positions in responsible DSR, such as the research integrators who help orchestrate the inter-, multi-, or transdisciplinary research projects and programs needed to address wicked problems (Hoffmann et al., 2022). This is not to say, however, that everyone else can also benefit from adopting a pragmatist worldview (Dewey, 1938b; Ulrich, 2006).

In terms of *perspectival knowing,* I argue that developing *reflexivity* provides an important opportunity for responsible DSR practitioners to enhance their perspective-taking capacities. While reflexivity is a contested concept that has been controversially discussed in the literature (e.g., Barad, 2014; Serra Undurraga, 2021), we use Stirling's (2006) helpful distinction between *reflectiveness* and *reflexivity* as a reference point to ground our understanding of the concept. Using a mirror as a metaphor for inquiry, reflectiveness can be understood as the "faithful reflection of all that lies in the field of view" (Stirling, 2006, p. 227), whereas reflexivity concerns the recognition of a subject who, when looking into the mirror, may also become a large part of the reflected image. Thus, reflexivity involves not only looking at the reflection itself, but also recognizing "the way[s] in which the attributes of the subject help condition [. . . the reflection] and how these representations themselves can help recondition the subject" (Stirling, 2006, p. 227). As such, we understand reflexivity to be intimately concerned with how we relate to the world and what this produces for ourselves and others (Serra Undurraga, 2021).

At the heart of reflexivity is the recognition that because of the plurality of uniquely different actors in the world and the inherent selectivity of all of our framings, there are multiple, potentially equally justified perspectives with which we must reckon (Stirling, 2006; Ulrich, 2006). This understanding requires us to engage in collaborative processes of critical engagement with the assumptions, knowledge, values, and practices that guide our work in order to collectively cocreate new, contextually appropriate ways of framing, knowing, organizing, and doing that move us closer to the successful resolution of problematic situations (Avelino et al., 2019; Popa et al., 2015). To put it bluntly, reflexive research cannot be done alone in an isolated ivory tower, but requires active participatory engagement and value co-creation with actors in the world. Only in interaction with the world can we begin to appreciate what our ways of relating produce for ourselves and others (Dewey, 1938a; Serra Undurraga, 2021).

Reflexivity can be nurtured and developed at two levels—the individual and the collective level. At the individual level, the focus is on developing self-reflexivity as a personal quality of a competent professional. Here, developing reflexivity is primarily about recognizing normative tensions in particular settings and cultivating the courage to proactively engage with them (e.g., Wittmayer & Schäpke, 2014). In practice, this requires being aware of one's own position in terms of time, place, background, and normativity, and how these factors may influence situational dynamics, as well as actively inviting other perspectives to round out the portfolio, so to speak.

At the collective level, the focus is on facilitating open-minded and open-ended processes of participatory inquiry that integrate the concerns and perspectives of relevant parties into all stages of the research (e.g., Popa et al., 2015). Here, developing reflexivity is closely related to efforts aimed at mainstreaming transdisciplinarity, which focus on supporting the management and design of more participatory and inclusive research projects and programs (e.g., Jahn et al., 2012; Popa et al., 2015). In practice, this means addressing the paradigmatic challenges of responsible DSR outlined in the previous sections.

With respect to *procedural knowing*, I join others in arguing that *sensitivity* is a core development opportunity for researchers who strive to be productive and effective. In this context, sensitivity refers to the ability to perceive relevant signals in the environment and develop contextually appropriate responses (e.g., Bryant, 2017; Glaser & Strauss, 1967). Two forms of sensitivity are generally distinguished in research settings—*theoretical sensitivity* and *methodological sensitivity* (Bryant, 2021).

Theoretical sensitivity as a term was first introduced by Glaser and Strauss (Glaser & Strauss, 1967, p. 46) to refer to the personal capacity of researchers "to have theoretical insight into [their] area of research, combined with an ability to make something of [their] insights" based on the "armamentarium of categories and hypotheses" that are built up through engagement with the research context and conditioned by the researchers' "personal and temperamental bent". As such, theoretical sensitivity concerns the development of researchers' capacities to function as useful research instruments, capable of detecting relevant signals from a variety of different inputs and integrating them into creative, contextually appropriate theoretical framings (see section on the ontological challenge) that can then be further developed, critiqued, and tested in inquiry (see section on the epistemological challenge). One aspect to guard against in this context, however, is the development of apophenia—an overly automatic form of pattern matching that leads to the spontaneous perception of connections between unrelated phenomena (Bryant, 2017, p. 247).

Building on the notion of theoretical sensitivity, Bryant (2017, p. 36) introduced the complementary term methodological sensitivity, which refers to "the skill or aptitude required by researchers in selecting, combining, and employing methods, techniques, and tools in actual research situations". It is similar to theoretical sensitivity in that both concepts concern the ability to detect relevant signals in the environment and to develop contextually appropriate responses; it differs from theoretical sensitivity in that it is concerned with the management and coordination of practices rather than the theorizing of conceptual structures.[16] It is a core competency for effective researchers because it is rarely, if ever, possible (or even desirable) to fully plan an inquiry in advance. Because the world is dynamic, diverse, complex, and constantly changing, inquiry must also be adaptive and responsive to the needs of the evolving situation (Andersen et al., 2022; Dewey, 1938b; Ulrich, 2003). Simply put, if there is no hope of designing the perfect project or program, the best we can do is to continually monitor our inquiries and use

[16]In a sense, theoretical sensitivity corresponds most closely to the aspects discussed as part of the ontological and epistemological challenges, while methodological sensitivity corresponds more closely to the axiological and methodological challenges.

the best methods, techniques, and tools at our disposal to pragmatically steer toward the successful resolution of problematic situations. Methodological sensitivity can be strengthened in several ways, including collaborating with and learning from other (more experienced) researchers; studying of research frameworks, methods, techniques, and tools; and, most fundamentally, learning by doing (Bryant, 2017; Dewey, 1938a; Schön, 1984). The framework presented in the section on the methodological challenge may prove useful in structuring such development efforts.

Finally, considering *propositional knowing*, I argue that enhancing *anticipation* through the development of anticipatory ways of thinking (Klein et al., 2010) is a key development opportunity and success criterion for responsible DSR (e.g., Jirotka et al., 2017; Stilgoe et al., 2013). Anticipatory thinking, defined by Klein et al. (2010, p. 235) as "the process of imagining how unexpected events may affect plans and practices", contrasts with the general human tendency to be overly optimistic when planning for the future (e.g., as evidenced by the prevalence of the planning fallacy; Buehler et al., 2010). It is a form of sensemaking (Weick et al., 2005) and macrocognitive function that emphasizes a future-oriented focus on anomalies, weak signals, and what can go wrong rather than what can be expected to go right, and, thus, can help us prepare for and protect against potential threats (Klein et al., 2010). It differs from prediction in that it is not just about guessing the most likely future states of the world, but about proactively preparing for all relevant scenarios, including those that are low-probability but potentially catastrophic (Klein et al., 2010).

According to Klein et al. (2010), three common forms of anticipatory thinking can be distinguished: *pattern matching*, *trajectory tracking*, and *convergence*. Pattern matching refers to the recognition of familiar situations and is closely related to a person's theoretical and methodological sensitivity, as discussed abvove. Trajectory tracking uses pattern matching, knowledge of associations between events, and extrapolation of trends to continuously monitor evolving situations. Convergence involves reasoning about the connections and interdependencies between different events and recognizing potentially relevant inferences they point to. Together, these forms of anticipatory thinking ideally culminate in the ongoing construction of useful, adaptive narratives that provide strategic guidance and mobilize responsive action even in complex, turbulent, and hard-to-predict environments (Klein et al., 2010).

For anticipatory thinking to be effective, there must be a responsive flow of action from the generation of anticipatory insights (i.e., the recognition that something needs urgent attention) to the implementation of appropriate responses (Klein et al., 2010; Stirling, 2006). Particularly in the context of larger organizations or societal and global governance institutions, this is not an easy task because flows of action can be stopped, redirected, or undermined at many points along the way. Some examples of barriers and obstacles include, but are not limited to, misaligned incentives, blocked or inefficient information flows, uncritical and conformist groupthink, or a mindless and dismissive general attitude (e.g., Ba et al., 2001; Klein, 2006; Klein et al., 2010; Stirling, 2006; Weick et al., 1999). How to overcome such barriers and effectively facilitate anticipatory thinking across scales—from the individual to our global societies—should be a central research question for responsible DSR. Simply put, given the accelerating pace of technological development, developing our capacity to anticipate and respond to potentially threatening developments will play a key role in determining the future trajectory of human civilization (e.g., Baum et al., 2019). It is my hope that the four paradigmatic challenges and associated recommendations I have developed in this paper can usefully support such efforts.

6 Discussion

The accelerating pace of development of increasingly life-altering digital technologies, such as data harvesting platforms (Zuboff, 2015), social media (Seger et al., 2020), or deep learning-based large language models (OpenAI, 2023), is putting increasing pressure on our societies. On the one hand, there is pressure to reap the potential benefits promised by digital technologies. On the other hand, there is a need to avoid the potentially catastrophic downsides that loom if we fail to anticipate the unintended or control the negative consequences of the digital transformation of our societies (e.g., Bostrom, 2013; Consilience Papers, 2022; Harari et al., 2023; Ord, 2020). As I have argued throughout this paper, IS research can play a key role in guiding our societies through this challenging situation, if we deliberately focus on finding responsible ways forward that help us carefully thread the needle.

This paper contributes to this goal by laying the groundwork for the emergent, participatory design (e.g., Bilandzic & Venable, 2011; Cavallo, 2000; Haj-Bolouri et al., 2016) of a responsible DSR ecosystem that can effectively facilitate this kind of work. On the one hand, the framework for responsible DSR provides a coherent and well-grounded conceptual basis for such development efforts. It complements previous research efforts in this direction by being not only *actionable* and *comprehensive* in scope, but also *consilient*, *flexible*, *open*, *parsimonious*, and specifically designed to facilitate *responsible* action and innovation, as summarized in Table 11. On the other hand, the four recommendations for the development of a responsible DSR ecosystem demonstrate the applicability and practical utility of the framework by considering its application in practice and outlining possible avenues for future research.

Table 11: Implementation of the Requirements for a Responsible DSR Framework.

Requirement	**Implementation**
Actionable	The framework is actionable because it distills the essence of responsible DSR practice into a set of four interdependent challenges, each of which can inform more effective and responsible action, as evidenced by the practical recommendations derived from it.
Comprehensive	The framework is comprehensive because it covers the four dimensions of research paradigms that have been recognized as most important for a holistic assessment: ontology, epistemology, axiology, and methodology.
Consilient	The framework is consilient because it facilitates the integration of different perspectives into a nuanced understanding of the world. For example, rather than perpetuating a strict division between behavioral science and design science, the framework recognizes the unity of all inquiry as a form of problem solving.
Flexible	The framework is flexible because it focuses only on the essence of responsible DSR practice, which is similar across application domains. No matter what problematic situation is being investigated, the four paradigmatic challenges will always be more or less relevant.

Requirement	Implementation
Open	The framework is open because it is consilient and flexible enough to be approached from different perspectives. It is also transparent in the sense that I have explained its development, illustrated it with empirical data, and outlined its implications. Moreover, the recommendations openly call for critical review and continued adaptation in the future.
Parsimonious	The framework is parsimonious in that it distills the complexity of responsible DSR practice into a manageable set of four interdependent challenges that together cover responsible DSR holistically.
Responsible	The framework supports responsible action and innovation by showing how they depend on the appropriate interplay of all four paradigmatic challenges, even across scales. Decisions in one area can affect all the others. A holistic perspective such as that presented in this framework is a useful starting point for implementing of a more systematic approach to ensuring responsible action and innovation.

6.1 Implications for Information Systems Theory

Three major implications of this paper for IS theory are worth highlighting: (1) it provides a novel perspective on paradigms in IS research; (2) it advances responsible DSR as a "super-methodology" for IS research; (3) it challenges the way in which we tend to draw paradigmatic boundaries in IS research.

First, at a very general level, the framework for responsible DSR offers a novel perspective on how to think about paradigms in IS research. Historically, there have been two main streams of research related to this topic.

On the one hand, there has been a stream of research focused on the role that paradigms should play at the level of the IS community as a whole. This was inspired by the Kuhnian insight that our most productive fields of research are underpinned by paradigms that provide the necessary conceptual structure to ensure that distributed work can productively complement and build upon each other (Kuhn, 2012). While some took this insight to suggest the need for an overarching paradigm for the IS field, others opposed this conclusion, arguing for a more pluralistic field that should not be tied to a singular paradigm (cf. Teo & Srivastava, 2007). Over time, the intensity of this debate subsided and the pluralist perspective prevailed, as evidenced by the broad intellectual diversity of IS research (e.g., Tarafdar et al., 2022) and the still ongoing calls for more systematic, strategic, inter-, multi-, or transdisciplinary programs of research (e.g., Gable, 2020; Nunamaker et al., 2017; Ram & Goes, 2021).

On the other hand, there is a tradition of using paradigmatic analyses to contrast different research approaches in terms of their basic assumptions and positions (e.g., Chen & Hirschheim, 2004; Iivari, 1991; Iivari, 2007). This stream of research dates back to the emergence of qualitative research as a new research approach that needed to be positioned against the dominant positivist approach (Morgan, 2007). Its core logic is to contrast and compare the competing research approaches in terms of their basic metaphysical assumptions about the nature of reality (i.e., ontology), truth (i.e., epistemology), value (i.e., axiology), and method (i.e., methodology), leading to the legitimation of multiple positions upon which researchers can build their work (Morgan, 2007). A key feature of this logic is its inherent tendency to limit the interoperability of research approaches, since one cannot be consistent and subscribe to multiple conflicting metaphysical positions at the same time (Morgan, 2007). More recently, as the demand for inter-, multi-, and transdisciplinary work has increased, this logic has come under criticism, with scholars calling for less emphasis on justifying research based on metaphysical positions and a more pragmatic and contextual focus on doing what can be shown to work in practice (e.g., Morgan, 2007).

This paper advances a third perspective on paradigms that combines aspects of both streams of research. Specifically, it recognizes the paradoxical role of paradigms in IS research and resolves this tension through a novel shift in perspective. While a paradigm is arguably necessary to facilitate a productive program of research (Kuhn, 2012), the way in which paradigms tend to be used to contrast research approaches in terms of the metaphysical positions actually limits our ability to productively conduct systematic, strategic, inter-, multi-, or transdisciplinary programs of research because it remains unclear how we can reasonably cross the metaphysical boundaries that we have established (Morgan, 2007).

This paper suggests that this paradox can be overcome if we define paradigms in a way that does not unduly limit the range of methodological approaches or theoretical perspectives that can be brought to bear on a research phenomenon. Nevertheless, they should be defined in such a way as to usefully guide action and comprehensively address the issues raised by potentially conflicting metaphysical positions. To achieve this, this paper defines the responsible DSR paradigm not in terms of legitimate metaphysical positions, as traditional paradigmatic analysis does, but by zooming in on and conceptualizing the challenges that the metaphysical positions aim to address in practical terms. Thus, this paper proposes a significant shift in perspective by defining a paradigm not in terms of specific positions, but in terms of the key practical, paradigmatic challenges that our work must face. This shift is inspired by a pragmatist understanding of research that is concerned with metaphysical positions only insofar as they make a difference in practice (James, 1922, pp. 43–46).

A key implication of this third perspective on paradigms for IS theory is that it provides a novel understanding of how to resolve the tension created by the paradoxical role of paradigms in IS research to date. An overarching IS research paradigm that can facilitate a productive, cumulative IS research tradition need not conflict with a pluralistic IS field if it is defined in a way that allows for a productive, open-minded, and open-ended integration of different types of work. The framework for responsible DSR illustrates how such a paradigm can be constructed. It is defined in terms of pragmatic inquiry that must address four paradigmatic challenges that point to key issues faced by researchers in practice. This definition encourages a creative, open-minded, and open-ended exploration of how these challenges can be addressed, rather than offering ready-made answers that do not take into account the specificities of a particular situation. It does not limit a priori what research phenomena can be investigated or what methods can be used, but it provides a basic conceptual overview of what practical challenges need to be addressed in order for a productive field of research to emerge. In this way, a paradigm can productively support cumulative work and resolve the tension around paradigms prevalent in IS research to date.

Second, an outcome of this shift in perspective has been a clarification of how DSR in IS can be adapted to serve as a "super-methodology" for systematic, strategic, inter-, multi-, or transdisciplinary programs of research (Nunamaker et al., 1990; Nunamaker et al., 2013). While previous research has recognized that all forms of research, even across scales, follow similar general problem-solving patterns (e.g., Dewey, 1938b; Farrell & Hooker, 2012; W. Kuechler et al., 2005; Nunamaker et al., 1990), with some notable exceptions (e.g., Ginige et al., 2018; Nunamaker et al., 2013; Nunamaker et al., 2017), little progress has been made in IS research to use this insight to develop a "super-methodology" that could systematically support the integration of our mostly distributed problem-solving efforts into a consilient, cumulative tradition. This paper addressed this gap by proposing a framework for responsible DSR that is specifically designed to serve as a "super-methodology" for IS research aimed at supporting and ensuring RI. A key feature of this framework is that it conceptualizes responsible DSR as a fallible, open-ended, discursive process that evolves iteratively over long periods of time and depends on an entire ecosystem of actors for its success. In doing so, the framework emphasizes the complex, collective, and cumulative nature of research and explicitly calls for more systematic coordination and integration efforts. It improves on the few previous works with a more comprehensive perspective.

A direct implication of this framework for IS theory is that it provides a better understanding of the key practical challenges that a "super-methodology" for IS research needs to address. It illustrates how the four paradigmatic challenges that it articulates are fundamental to all forms of a research and are relevant across scales. It shows how they can provide a common foundation, but also open up new perspectives on the possibilities and limitations of responsible DSR. In this way, the framework promotes a comprehensive theoretical understanding that can inspire and guide future research towards a more systematic engagement with the question of how to develop and organize a responsible DSR ecosystem. This responds to recent calls for a more systematic study and development of DSR and RI ecosystems (e.g., Morana et al., 2018; Stahl, 2022; Timmermans et al., 2020).

Third, the framework for responsible DSR also challenges the way we tend to draw paradigmatic boundaries in IS research. Much as qualitative research approaches were positioned against the more established positivist research approaches to gain legitimacy in their early years (Morgan, 2007), DSR was introduced as a new research paradigm, complementary to but ultimately incommensurable with the more established BSR paradigm (e.g., Hevner et al., 2004; March & Smith, 1995). However, this strategic positioning quickly became questionable and difficult to sustain in practice when it became clear that work in each paradigm not only regularly incorporates methods, but can also produce results that we would generally attribute to the domain of the other (Baskerville et al., 2015; Herwix & Rosenkranz, 2018). Nevertheless, this insight has yet to be fully assimilated on the theoretical side, with a recent framework for the integration of BSR and DSR still maintaining the traditional juxtaposition between the two (see Seidel & Watson, 2020). The framework for responsible DSR finally offers an alternative view in which responsible DSR takes the position of a "super-methodology" that facilitates the systematic coordination and integration of all forms of IS research into strategic, inter-, multi-, or transdisciplinary programs of research.

An important implication of this alternative view for IS theory is that it presents a well-founded challenge to the received wisdom in IS research that has considerable critical potential and thus can inform future theorizing. While it may be a tall order to challenge the now deeply ingrained way in which we draw paradigmatic boundaries between BSR and DSR in IS research, I argue that the view presented in this paper offers a promising and less problematic alternative that could support more systematic cooperation within the IS research community and significantly advance its collective value co-creation capacity. Although changing such deeply entrenched, taken-for-granted assumptions may be difficult and costly, the expected benefits on the other side could make such efforts worthwhile. At the very least, the availability of an alternative view should prompt us to critically reflect on our own taken-for-granted assumptions, and spur future research to more systematically investigate the costs and benefits of the paradigmatic boundaries we reinforce with our work.

6.2 Implications for Information Systems Practice

I highlight two main implications of this paper for IS practice: (1) the vision presented here can serve as a useful boundary object that can engage and bring together the parties relevant to the development of a responsible DSR ecosystem, and (2) the recommendations provide more specific guidance that can inspire the exaptation and reorientation of existing institutions and practices to support the realization of this vision.

First, the paper presents a vision for a responsible DSR ecosystem through a well-grounded and comprehensive framework and a set of recommendations for its implementation. As discussed by Winter & Butler (2011), such work can be highly beneficial to the IS community if it can attract and sustain the interest of parties relevant to its success, such as research funders, IS practitioners, researchers from other fields, and policy makers. A key factor in this regard is whether it has the capacity to serve as a boundary object (Star & Griesemer, 1989), capable of bridging the different social worlds that each of us inhabits. In general, this can be the case when it "identif[ies] 'lowest common denominators,' critical points of agreement, or shared surface referents [...] to provide a sufficient platform for cooperative action, [...] without requiring the individuals involved to abandon the distinctive perspectives, positions, and practices of their 'base' social world" (Winter & Butler, 2011, p. 103). The vision presented in this paper aims to do just that.

For example, the framework for responsible DSR provides a lowest common denominator for responsible IS research practice by capturing only the essential paradigmatic challenges faced in practice. It does so without asking individuals to abandon their unique perspectives, but rather recognises the inherent selectivity of any perspective and thus calls us to engage in critical discourse with each other's understandings. In this way, it can bridge the different perspectives on IS research and systematically facilitate more strategic, inter-, multi-, or transdisciplinary programs of research. In addition, the associated recommendations for developing a responsible DSR ecosystem paint a broader picture of the vision, making it accessible and relevant to parties outside the IS field, such as research funders, IS practitioners, researchers from other fields, and policy makers. Taken together, this suggests that the vision presented in this paper can indeed serve as a useful boundary object that can rally relevant parties in the common pursuit of a more desirable and responsible future.

However, it is important to recognize that an aspirational vision that can serve as a boundary object is a necessary but not a sufficient condition for the successful realization of a more desirable future. While we need to imagine the future if we are to intentionally shape it (e.g., Gaziulusoy & Ryan, 2017; Gümüsay & Reinecke, 2022; Mahfoud, 2021; Osterwalder & Pigneur, 2010), ultimately our future remains inherently unpredictable given the fundamental uncertainty at play in social systems (e.g., Beckert, 2013; Jonas, 1984; Rittel & Webber, 1973; Ulrich, 1994). Therefore, the vision presented in this paper should be seen as no more and no less than a critical contribution that provides a common starting point for individual reflection and further discourse on how we can come together to leverage the strengths and overcome the weaknesses of the IS community to condition more responsible real-world action in practice. It will take a sustained effort by many different parties to carry this vision forward and negotiate its implementation in practice.

Second, the four recommendations for implementation presented in this paper provide specific guidance that can inspire the exaptation and reorientation of existing institutions and practices to support the realization of this vision. For example, the MLDM offers a novel vision that reimagines how we might structure our research efforts and discourse if we were determined to come together collectively as a research ecosystem and act responsibly at every stage of research. This includes the specific proposal to exapt the existing peer review institutions and practices for a more comprehensive feedback system that helps to coordinate and streamline research efforts by intervening much earlier in the research process. The goal of this system would be to help the IS community gain the collective capacity to continuously identify and develop its key priorities, and to address them productively by redirecting its attention and resources. It is clear that such a system could be very beneficial to the IS community if it were to enhance our collective ability to address larger and more significant problems by forming "substantial communities of inquiry in a timely fashion" (Winter & Butler, 2011, p. 107). This would not only help us increase the relevance, visibility, and legitimacy of our work, but also our collective productivity by avoiding unnecessary, what Rai (2017b) calls, Type III errors: addressing research problems that do not matter. However, it would undoubtedly be difficult to implement because it would require multiple parties to come together, align their incentives, and change their behaviors and practices at the same time.

Therefore, the service-dominant logic framework and the CDPs introduced as part of the third recommendation provide useful perspectives and tools that have demonstrated the ability to analyze and support even very difficult and complex cultural change processes toward greater cooperation and value co-creation (e.g., Lusch & Vargo, 2014b; Vargo et al., 2020; D. S. Wilson et al., 2023; D. S. Wilson et al., 2013; D. S. Wilson et al., 2020). A key takeaway from this literature is to actively identify, acknowledge, and engage with the diverse interests and incentives present in a complex social system such as the IS community. Rather than assuming a good understanding of the situation, long-term success requires actually making the effort to consult all concerned parties on a regular basis (Ostrom, 2010; Ulrich, 1994). The CDPs provide a holistic framework for organizing such engagement processes and can thus inform their practical implementation.

Overall, it should be clear from these discussions that nothing offered in this paper can be seen as a panacea, or even as a clear-cut proposal for how to move forward in practice. Life in our digital age simply defies such simple solutions (Jirotka et al., 2017; Rittel & Webber, 1973). The main goal of this paper, therefore, is simply to take a first step by challenging taken-for-granted assumptions about how IS research should be conducted by presenting a well-grounded, desirable, alternative vision of how things could, or even *should*, be. This means that the actual implications of this paper for IS practice are difficult, if not impossible, to predict. It must be up to each individual and the IS community as a whole to decide how to respond.

6.3 Implications for Policy Making

This paper can inform policy making in two ways: (1) the MLDM provides a conceptual starting point for regulating high-risk technology development processes; (2) the paper as a whole makes the case for recognizing and funding an open responsible DSR ecosystem as a coordination and support system for responsible DSR projects and programs.

First, the MLDM provides a basic conceptual template for how to regulate high-risk technology development processes with the goal of ensuring RI. While it can certainly only be a starting point for the development of such regulation, it makes two key recommendations that such efforts should consider: *multi-layered* and *stage-based* development.

Multi-layered development highlights how the focus of a development process naturally expands over time as it moves closer to a resolution: from a dedicated focus on identifying a relevant problem at the beginning, to a broader focus on responsibly deploying an identified solution to the public at the end of the process. Thinking of the development process as multi-layered helps to recognize that a variety of expertise is needed to ensure that development is pursued in a responsible manner, and emphasizes that responsible development should ideally be inter-, multi-, or transdisciplinary (Kroeze et al., 2019). In particular, it is not enough to consider each layer in isolation. Even when addressing seemingly trivial problems, but especially when developing high-risk technologies, responsibility requires that we always consider the impact of our work on the whole system (Ulrich, 1994). Community-based discourse at each layer can help ensure that a range of perspectives are available for consideration during development.

Stage-based development proposes staging the development process based on meeting certain criteria that demonstrate responsible progress and behavior. This can be seen as a counterpoint to Mark Zuckerberg's internal motto for Meta: "move fast and break things" (Blodget, 2009). Rather than naively prioritizing speed at all costs, the goal is to ensure productivity in a responsible way by identifying problems as early in the development process as possible. This means investing significant resources in understanding the problem situation we are in, looking at it from different perspectives, including its relationship to the whole system, to anticipate and address potential failure modes, externalities, and other unintended consequences. To ensure this, mandatory stage-gates in the development process could ensure that minimum quality standards are upheld.

For example, considering high-risk development processes such as those involving advanced AI technologies, before any prototypes are allowed to be developed, it could be mandated that the development team must be interdisciplinary and be given the time and resources to get up to speed on the latest developments in AI safety research to ensure that they understand the risk landscape they are working with, as well as any other topics that may be relevant to the problem they are working on. In practice, this stage-gate could take the form of an open peer review process in which the team seeking to move forward must demonstrate that it is following good practices and has the relevant expertise to do the work it is seeking to do, for example through appropriate certifications and qualifications. Later stage-gates in the development process could be treated in a similar way. This should lead to a more accountable, but still decentralized and self-organizing, feedback system for development processes. One that does not impose a particular development direction from the outside, but still ensures that relevant expertise is available and good practices are followed, while also creating transparency about who is working on what and at what stage. Adherence to stage-gates could be enforced by laws requiring the open documentation of all high-risk development processes, and allowing all citizens to sue high-risk development efforts that cannot demonstrate that they have followed appropriate practices and standards.

Needless to say, such a feedback system would be controversial and would not, by itself, solve all the problems associated with regulating high-risk development processes. For example, the above discussion does not yet address the question of how to organize and fund such a comprehensive feedback system in a way that maintains both safety and productivity. Nevertheless, it provides two key points for a proposal that can stimulate critical reflection on the currently largely unregulated field of digital technology development. In doing so, it can ideally serve as a critical provocation that stimulates the development of more responsible policies.

Second, this paper as a whole argues for the recognition and funding of an open responsible DSR ecosystem as a coordination and support system for responsible DSR projects and programs. In particular, it recognizes that all research projects and programs are always collective endeavors that relate to and depend on an active community of inquiry that acts as a jury to review and evaluate their results (Pardales & Girod, 2006). As such, the value co-created by research efforts is inherently constrained by the environment in which they operate. This requires us to consider the whole system in which responsible DSR projects and programs are embedded. In practice, this means ensuring that an appropriate infrastructure is in place to coordinate and support responsible DSR projects and programs.

Taking this insight to heart implies the need for policy makers to provide appropriate funding schemes that can support the development and maintenance of the necessary infrastructure. While the nature of appropriate funding schemes is certainly debatable, I highlight the possibility of recognizing and funding research ecosystems directly. For example, in the wake of the Covid-19 pandemic, the German Federal Ministry of Education and Research funded the German Network of University Medicine (NUM), a collaboration of all university hospitals in Germany, to establish and maintain the necessary infrastructure to support high-quality research on Covid-19 (Schons et al., 2022). NUM activities include projects such as the German National Pandemic Cohort Network (NAPKON), which was established to exapt existing and create new institutions and practices to collect, process, analyze, and manage the data for the most comprehensive Covid-19 cohort in Germany (Schons et al., 2022). The rapid launch of such a large-scale research effort was only possible because funding was made available not only for individual projects, but also for the coordination and support system behind these projects. Without NUM, a large-scale effort like NAPKON would have been much more difficult to set up. Thus, investments in infrastructure that supports a research ecosystem can be viewed as an important, strategic enabler for high-impact research (Nunamaker et al., 2017; Winter & Butler, 2011).

This means that policy makers have an opportunity to make a significant contribution to the development of an open, responsible DSR ecosystem if they design funding schemes that help the entire ecosystem thrive. This paper presents two main sources of inspiration for such efforts. On the one hand, the MLDM argues for the need to exapt existing institutions and practices to create a comprehensive feedback system that can help coordinate and support work in the ecosystem more efficiently and responsibly. On the other hand, the four development opportunities identified as part of the fourth recommendation of this paper suggest a set of orientations, skills, and capabilities whose development should be supported in order to enhance the overall adaptability, quality, productivity, and impact of the responsible DSR ecosystem. While these two suggestions certainly do not constitute a comprehensive policy proposal, taken together they provide a reasonable and informed starting point for policy makers seeking to design policies and funding schemes to support the development of a responsible DSR ecosystem.

6.4 Limitations

It is important to acknowledge the major limitations of this work. As argued throughout this paper, all framings are inherently selective and must therefore be viewed as incomplete and even potentially misleading. While I have tried to be transparent about the boundaries drawn in this work and to illustrate their usefulness for framing responsible DSR, the findings presented are still nascent and require more significant engagement and critique from a broader set of actors. While I advocate for an emergent, participatory design approach to developing a responsible DSR ecosystem, the findings of this paper are not the result of such a participatory process. Although significant input from the DSR community has been incorporated into the project, the findings are ultimately generated through a traditional, non-participatory, multi-grounded theorizing process involving only a small research team. This means that the findings cannot be taken as representative of the perspectives of the IS research community at large, but rather must be seen as the personal perspective of the author. As such, this paper should only serve as a starting point for further critical discourse and more emergent, participatory design efforts in the future.

In addition, this relatively nascent state of the findings makes it difficult, if not impossible, to objectively assess the quality and contribution of this work. From a pragmatist perspective, what is needed to reach a final judgment is a comprehensive transformation of the problematic situation into one that is more or less resolved (Dewey, 1938b). Until this point of resolution is reached, all findings must be regarded as highly tentative (Dewey, 1941). Given the ambition of this paper, however, it is not possible to claim to have reached a state of resolution. Quite the opposite, in fact. If this paper is well received and taken seriously, it is likely to generate a good deal of controversy and critical discussion about the nature of responsible DSR and what it should mean for the identity and direction of the IS research community.

This situation leaves us with the question of how to judge the quality of this work, if not by its demonstrated ability to resolve a problematic situation? I suggest that the answer lies in assessing its critical potential. As the complexity of our world increases in the digital age, comprehensive resolutions to problematic situations become more elusive. As Ulrich (1994) has pointed out, in such situations, contributions that help us see problematic situations from a new perspective and in a different light often turn out to be the most relevant ones in retrospect. In this sense, meaningful and reasonable critique can be seen as the engine that drives progress toward eventual resolution, and a useful ideal against which to measure this work. So the question becomes: Does this paper make a reasonable case for a new and meaningful way of looking at IS research that can expand and inform our thinking? I certainly hope so, and invite the reader to make up his or her own mind.

7 Conclusion

In conclusion, this paper has highlighted the growing importance of IS research in addressing the societal challenges posed by the rapid development of digital technologies. The need to manage the benefits and potentially catastrophic downsides of the ongoing digital transformation has become a central concern for our societies. While the understanding of how to contribute to this imperative is still in its early stages, recent developments, such as the statement by many eminent AI researchers and other notable figures that "mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war" (Center for AI Safety, 2023), underscore the urgency of the situation.

To address this problematic situation, this paper proposes the development of a responsible DSR ecosystem as a promising way forward. This responsible DSR ecosystem should integrate insights from various fields and disciplines to ensure that we can responsibly address the grand challenges facing our societies. As a unifying foundation for this responsible DSR ecosystem, this paper has identified four fundamental paradigmatic challenges that such an ecosystem must address: ontological, epistemological, axiological, and methodological. These challenges provide a holistic framework that is both theoretically grounded and practically viable. It can serve as a boundary object to bring together relevant parties interested in creating a responsible DSR ecosystem. In addition, the paper offers four specific recommendations to support the development of a responsible DSR ecosystem based on the framework.

In summary, this paper calls on the IS research community to embrace responsible DSR and contribute to its mission of responsibly addressing the societal challenges posed by the digital transformation. By critically engaging with the proposed paradigmatic framework and the provided recommendations, IS researchers can play a pivotal role in shaping the future of a responsible DSR ecosystem. Such efforts are essential to ensure that the digital technologies we develop remain beneficial to humanity and do not inadvertently increase potentially catastrophic risks. If long-term societal prosperity and well-being are our research questions, responsible DSR is the inquiry to find the answers.

Acknowledgements

I sincerely thank all of the individuals who contributed to this research project and the development of this paper. Their input, support and insights were invaluable and without them this study would not have been possible. In particular, I would like to thank the interviewees for sharing their experiences and expertise, which challenged my thinking and enriched my understanding of the subject matter. I am also grateful to all the researchers who provided feedback that improved the clarity, rigor, and quality of this work. Overall, I am sincerely grateful to everyone involved, including those inadvertently omitted, for shaping the outcome of this research.

Online Appendix

Herwix, A. (2023). *Threading the Needle in the Digital Age: Four Paradigmatic Challenges for Responsible Design Science Research (Online Appendix)*. OSF. https://doi.org/10.17605/OSF.IO/75VDH.

References

Ackoff, R. L. (1994). Systems thinking and thinking systems. *System Dynamics Review*, *10*(2-3), 175–188. https://doi.org/10.1002/sdr.4260100206

Alter, S. (2012). Long Live Design Science Research! .... And Remind Me again about Whether It Is a New Research Paradigm or a Rationale of Last Resort for Worthwhile Research that Doesn't Fit under Any Other Umbrella. *ICIS 2012 Proceedings*, 6.

Alter, S. (2013). Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future. *Journal of the Association for Information Systems*, *14*(2), 72–121. https://doi.org/10.17705/1jais.00323

Andersen, B. P., Miller, M., & Vervaeke, J. (2022). Predictive processing and relevance realization: Exploring convergent solutions to the frame problem. *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-022-09850-6

Ashby, W. R. (1968). Variety, Constraint, and the Law of Requisite Variety. *Systems research for behavioral science*. Routledge.

Askell, A., Brundage, M., & Hadfield, G. (2019). The Role of Cooperation in Responsible AI Development. *arXiv:1907.04534 [cs]*. Retrieved October 2, 2019, from http://arxiv.org/abs/1907.04534

Association for Information Systems. (2023). *Senior Scholars' List of Premier Journals*. Retrieved May 2, 2023, from https://aisnet.org/page/SeniorScholarListofPremierJournals

Avdiji, H., Elikan, D., Missonier, S., & Pigneur, Y. (2020). A Design Theory for Visual Inquiry Tools. *Journal of the Association for Information Systems*, *22*(2), 247–265. https://doi.org/10.17705/1jais.00491

Avelino, F., Wittmayer, J. M., Pel, B., Weaver, P., Dumitru, A., Haxeltine, A., Kemp, R., Jørgensen, M. S., Bauler, T., Ruijsink, S., & O'Riordan, T. (2019). Transformative social innovation and (dis)empowerment. *Technological Forecasting and Social Change*, *145*, 195–206. https://doi.org/10.1016/j.techfore.2017.05.002

Baker, A. (2022). Simplicity. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2022). Metaphysics Research Lab, Stanford University. Retrieved April 15, 2023, from https://plato.stanford.edu/archives/sum2022/entries/simplicity/

Barad, K. (2014). Diffracting Diffraction: Cutting Together-Apart. *Parallax*, *20*(3), 168–187. https://doi.org/10.1080/13534645.2014.927623

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design Science Research Contributions: Finding a Balance between Artifact and Theory. *Journal of the Association for Information Systems*, *19*(5), 358–376. https://doi.org/10.17705/1jais.00495

Baskerville, R., Kaul, M., & Storey, V. C. (2015). Genres of Inquiry in Design-Science Research: Justification and Evaluation of Knowledge Production. *MIS Quarterly*, *39*(3), 541–564.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*(3), 1–52. https://doi.org/10.1145/1541880.1541883

Baum, S. D., Armstrong, S., Ekenstedt, T., Häggström, O., Hanson, R., Kuhlemann, K., Maas, M. M., Miller, J. D., Salmela, M., Sandberg, A., Sotala, K., Torres, P., Turchin, A., & Yampolskiy, R. V. (2019). Long-term trajectories of human civilization. *foresight*, *21*(1), 53–83. https://doi.org/10.1108/FS-04-2018-0037

Beckert, J. (2013). Imagined futures: Fictional expectations in the economy. *Theory and Society*, *42*(3), 219–240. https://doi.org/10.1007/s11186-013-9191-2

Benke, I., Feine, J., Venable, J., & Maedche, A. (2020). On Implementing Ethical Principles in Design Science Research. *AIS Transactions on Human-Computer Interaction*, *12*(4), 206–227. https://doi.org/10.17705/1thci.00136

Berenbaum, M. R. (2023). On peer review—then, now, and soon to be? *Proceedings of the National Academy of Sciences*, *120*(11), e2302593120. https://doi.org/10.1073/pnas.2302593120

Bilandzic, M., & Venable, J. (2011). Towards participatory action design research: Adapting action research and design science research methods for urban informatics. *Journal of Community Informatics*, *7*(3). Retrieved May 20, 2023, from https://espace.curtin.edu.au/handle/20.500.11937/49022

Blodget, H. (2009). Mark Zuckerberg On Innovation. Retrieved December 10, 2023, from https://www.businessinsider.com/mark-zuckerberg-innovation-2009-10

Blok, V. (Ed.). (2023). *Putting Responsible Research and Innovation into Practice: A Multi-Stakeholder Approach* (Vol. 40). Springer International Publishing. https://doi.org/10.1007/978-3-031-14710-4

Blok, V., & Lemmens, P. (2015). The Emerging Concept of Responsible Innovation. Three Reasons Why It Is Questionable and Calls for a Radical Transformation of the Concept of Innovation. In B.-J. Koops, I. Oosterlaken, H. Romijn, T. Swierstra, & J. van den Hoven (Eds.), *Responsible innovation 2: Concepts, approaches, and applications* (pp. 19–35). Springer International Publishing. https://doi.org/10.1007/978-3-319-17308-5_2

Boh, W., Constantinides, P., Padmanabhan, B., & Viswanathan, S. (2023). Building Digital Resilience against Major Shocks. *Management Information Systems Quarterly*, *47*(1), 343–360.

Bostrom, N. (2013). Existential Risk Prevention as Global Priority. *Global Policy*, *4*(1), 15–31. https://doi.org/10.1111/1758-5899.12002

Bostrom, N., & Cirkovic, M. M. (2011, September 29). *Global Catastrophic Risks*. OUP Oxford.

Bostrom, N., Douglas, T., & Sandberg, A. (2016). The Unilateralist's Curse and the Case for a Principle of Conformity. *Social Epistemology*, *30*(4), 350–371. https://doi.org/10.1080/02691728.2015.1108373

Briggs, R. O., & Schwabe, G. (2011). On Expanding the Scope of Design Science in IS Research. In H. Jain, A. P. Sinha, & P. Vitharana (Eds.), *Service-oriented perspectives in design science research* (pp. 92–106). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20633-7_7

Brown, M. J. (2012). John Dewey's Logic of Science. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, *2*(2), 258–306. https://doi.org/10.1086/666843

Bryant, A. (2017). *Grounded Theory and Grounded Theorizing: Pragmatism in Research Practice*. Oxford University Press.

Bryant, A. (2021, February 16). An Overview of Grounded Theory Aka The Grounded Theory Method (GTM). *Handbook of Qualitative Research Methodologies in Workplace Contexts* (pp. 5–29). Edward Elgar Publishing. Retrieved May 4, 2023, from https://www.elgaronline.com/display/edcoll/9781789904338/9781789904338.00007.xml

Brynjolfsson, E., & McAfee, A. (2012, January 23). *Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*. Digital Frontier Press.

Buehler, R., Griffin, D., & Peetz, J. (2010, January 1). Chapter One - The Planning Fallacy: Cognitive, Motivational, and Social Origins. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (pp. 1–62). Academic Press. https://doi.org/10.1016/S0065-2601(10)43001-4

Burton-Jones, A., Boh, W., Oborn, E., & Padmanabhan, B. (2021). Editor's Comments: Advancing Research Transparency at MIS Quarterly: A Pluralistic Approach. *MIS Quarterly*, *45*(2), iii–xviii.

Carlile, P. R. (2002). A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development. *Organization Science*, *13*(4), 442–455. https://doi.org/10.1287/orsc.13.4.442.2953

Carlile, P. R. (2004). Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries. *Organization Science*, *15*(5), 555–568. https://doi.org/10.1287/orsc.1040.0094

Castells, M. (2011). *The Rise of the Network Society* (2. ed. with a new preface, [reprint]). Wiley-Blackwell.

Cavallo, D. (2000). Emergent Design and learning environments: Building on indigenous knowledge. *IBM Systems Journal*, *39*(3.4), 768–781. https://doi.org/10.1147/sj.393.0768

Center for AI Safety. (2023). *Statement on AI Risk*. Retrieved May 30, 2023, from https://www.safe.ai/statement-on-ai-risk

Chalmers, A. F. (2013). *What Is This Thing Called Science* (4th ed). University of Queensland Press.

Chambers, C. D., & Tzavella, L. (2020, February 10). *The past, present, and future of Registered Reports*. MetaArXiv. https://doi.org/10.31222/osf.io/43298

Chen, W., & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information Systems Journal*, *14*(3), 197–235. https://doi.org/10.1111/j.1365-2575.2004.00173.x

Churchman, C. W. (1967). Guest Editorial: Wicked Problems. *Management Science*, *14*(4), B-141-B–146. https://doi.org/10.1287/mnsc.14.4.B141

Collingridge, D. (1980). *The social control of technology*. Pinter.

Consilience Papers. (2022, June 26). Technology is Not Values Neutral: Ending the Reign of Nihilistic Design. Retrieved January 30, 2023, from https://consilienceproject.org/technology-is-not-values-neutral/

Cook, S. D. N., & Brown, J. S. (1999). Bridging Epistemologies: The Generative Dance Between Organizational Knowledge and Organizational Knowing. *Organization Science*, *10*(4), 381–400. https://doi.org/10.1287/orsc.10.4.381

Córdoba, J.-R., & Midgley, G. (2008). Beyond organisational agendas: Using boundary critique to facilitate the inclusion of societal concerns in information systems planning. *European Journal of Information Systems*, *17*(2), 125–142. https://doi.org/10.1057/ejis.2008.4

Davison, R. M., Majchrzak, A., Hardin, A., & Ravishankar, M.-N. (2023). Special issue on responsible IS research for a better world. *Information Systems Journal*, *33*(1), 1–7. https://doi.org/10.1111/isj.12405

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The Digital Platform: A Research Agenda. *Journal of Information Technology*, *33*(2), 124–135. https://doi.org/10.1057/s41265-016-0033-3 de Reuver, M., van Wynsberghe, A., Janssen, M., & van de Poel, I. (2020). Digital platforms and responsible innovation: Expanding value sensitive design to overcome ontological uncertainty. *Ethics and Information Technology*, *22*(3), 257–267. https://doi.org/10.1007/s10676-020-09537-z

Desouza, K. C., & Dawson, G. S. (2023). Doing strategic information systems research for public value. *The Journal of Strategic Information Systems*, *32*(4), 101805. https://doi.org/10.1016/j.jsis.2023.101805

Dewey, J. (1938a). *Experience and education* (Touchstone Edition). Touchstone.

Dewey, J. (1938b). *Logic: The Theory of Inquiry*. Henry Holt; Company, INC.

Dewey, J. (1941). Propositions, Warranted Assertibility, and Truth. *The Journal of Philosophy*, *38*(7), 169–186. https://doi.org/10.2307/2017978

Dewey, J. (2007). The Development of American Pragmatism. *Scientiae Studia*, *5*(2), 227–243.

Douglas, H. E. (2009). *Science, policy, and the value-free ideal* (1st). University of Pittsburgh Press.

Drechsler, A., & Hevner, A. (2022). Knowledge Paths in Design Science Research. *Foundations and Trends® in Information Systems*, *6*(3), 171–243. https://doi.org/10.1561/2900000028

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). “so what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Farrell, R., & Hooker, C. (2012). The Simon–Kroes model of technical artifacts and the distinction between science and design. *Design Studies*, *33*(5), 480–495. https://doi.org/10.1016/j.destud.2012.05.001

Farrell, R., & Hooker, C. (2013). Design, science and wicked problems. *Design Studies*, *34*(6), 681–705. https://doi.org/10.1016/j.destud.2013.05.001

Feyerabend, P. (1993). *Against method* (3rd ed). Verso.

Fleischmann, K. R. (2006). Boundary Objects with Agency: A Method for Studying the Design–Use Interface. *The Information Society*, *22*(2), 77–87. https://doi.org/10.1080/01972240600567188

Flew, A. (1979). Golden Rule.

Gable, G. G. (2020). Viewpoint: Information systems research strategy. *The Journal of Strategic Information Systems*, *29*(2), 101620. https://doi.org/10.1016/j.jsis.2020.101620

Galliers, R. (2003). Change as Crisis or Growth? Toward a Trans-disciplinary View of Information Systems as a Field of Study: A Response to Benbasat and Zmud’s Call for Returning to the IT Artifact. *Journal of the Association for Information Systems*, *4*(1), 337–352. https://doi.org/10.17705/1jais.00040

Gaziulusoy, A. İ., & Ryan, C. (2017). Roles of design in sustainability transitions projects: A case study of Visions and Pathways 2040 project from Australia. *Journal of Cleaner Production*, *162*, 1297–1307. https://doi.org/10.1016/j.jclepro.2017.06.122

Genus, A., & Stirling, A. (2018). Collingridge and the dilemma of control: Towards responsible and accountable innovation. *Research Policy*, *47*(1), 61–69. https://doi.org/10.1016/j.respol.2017.09.012

Germonprez, M., Hovorka, D., & Gal, U. (2011). Secondary Design: A Case of Behavioral Design Science Research. *Journal of the Association for Information Systems*, *12*(10), 662–683.

Gholami, R., Watson, R., Hasan, H., Molla, A., & Bjorn-Andersen, N. (2016). Information Systems Solutions for Environmental Sustainability: How Can We Do More? *Journal of the Association for Information Systems*, *17*(8), 521–536. https://doi.org/10.17705/1jais.00435

Giampietro, M., & Mayumi, K. (2018). Unraveling the Complexity of the Jevons Paradox: The Link Between Innovation, Efficiency, and Sustainability. *Frontiers in Energy Research*, *6*. Retrieved September 13, 2023, from https://www.frontiersin.org/articles/10.3389/fenrg.2018.00026

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. SAGE Publications.

Ginige, T., De Silva, L., Walisadeera, A., & Ginige, A. (2018). Extending DSR with Sub Cycles to Develop a Digital Knowledge Ecosystem for Coordinating Agriculture Domain in Developing Countries. In S. Chatterjee, K. Dutta, & R. P. Sundarraj (Eds.), *Designing for a digital and globalized world* (pp. 268–282). Springer International Publishing.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative research*. Aldine Transaction.

Godin, B. (2014). *Innovation contested: The idea of innovation over the centuries*. Routledge.

Goldkuhl, G., & Cronholm, S. (2010). Adding Theoretical Grounding to Grounded Theory: Toward Multi-Grounded Theory. *International Journal of Qualitative Methods*, *9*(2), 187–205. https://doi.org/10.1177/160940691000900205

Goldkuhl, G., & Cronholm, S. (2018). Reflection/commentary on a Past Article: "Adding Theoretical Grounding to Grounded Theory: Toward Multi-Grounded Theory". *International Journal of Qualitative Methods*, *17*(1). https://doi.org/10.1177/1609406918795540

Goldkuhl, G., & Karlsson, F. (2020). Method Engineering as Design Science. *Journal of the Association for Information Systems*, *21*(5). https://doi.org/10.17705/1jais.00636

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *32*(2), 337–355.

Gümüsay, A. A., & Reinecke, J. (2022). Researching for Desirable Futures: From Real Utopias to Imagining Alternatives. *Journal of Management Studies*, *59*(1), 236–242. https://doi.org/10.1111/joms.12709

Haj-Bolouri, A., Bernhardsson, L., & Rossi, M. (2016). Padre: A Method for Participatory Action Design Research. *Tackling Society's Grand Challenges with Design Science*, 19–36. https://doi.org/10.1007/978-3-319-39294-3_2

Harari, Y., Harris, T., & Raskin, A. (2023). Opinion — You Can Have the Blue Pill or the Red Pill, and We're Out of Blue Pills [newspaper]. *The New York Times*. Retrieved April 16, 2023, from https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html

Hartmann, T. (2012). Wicked problems and clumsy solutions: Planning as expectation management. *Planning Theory*, *11*(3), 242–256. https://doi.org/10.1177/1473095212440427

Hayes, S. C., Hayes, L. J., & Reese, H. W. (1988). Finding the Philosophical Core: A Review of Stephen C. Pepper's World Hypotheses: A Study in Evidence. *Journal of the Experimental Analysis of Behavior*, *50*(1), 97–111.

Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. *Electronic Markets*, *30*(1), 87–98. https://doi.org/10.1007/s12525-019-00377-4

Herwix, A., & Rosenkranz, C. (2018). Making Sense of Design Science in Information Systems Research: Insights from a Systematic Literature Review. In S. Chatterjee, K. Dutta, & R. P. Sundarraj (Eds.), *Designing for a digital and globalized world* (pp. 51–66). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-91800-6_4

Herwix, A., & Rosenkranz, C. (2019). A Multi-Perspective Framework for the Investigation of Tool Support for Design Science Research. *Proceedings of the 27th European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2019_rp/164

Herwix, A., Rossi, M., Purao, S., Haj-Bolouri, A., Tremblay, M. C., & Gregor, S. (2022). Ethics in Information Systems and Design Science Research: Five Perspectives. *Communications of the Association for Information Systems*, *50*(1), 589–616. https://doi.org/10.17705/1CAIS.05028

Hevner, A. R. (2020). *Daring to Do Good Design Science Research*. Retrieved February 27, 2023, from https://research.nii.ac.jp/decor/DECOR2020program.html

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science In Information Systems Research. *MIS Quarterly*, *28*(1), 75–105.

Hoffmann, S., Deutsch, L., Klein, J. T., & O'Rourke, M. (2022). Integrate the integrators! A call for establishing academic careers for integration experts. *Humanities and Social Sciences Communications*, *9*(1), 1–10. https://doi.org/10.1057/s41599-022-01138-z

Hooker, C. (2017). A proposed universal model of problem solving for design, science and cognate fields. *New Ideas in Psychology*, *47*, 41–48. https://doi.org/10.1016/j.newideapsych.2017.05.001

Hyde, A., Pattinson, D., & Shannon, P. (2022). Designing for Emergent Workflow Cultures: eLife, PRC, and Kotahi. *Commonplace*. https://doi.org/10.21428/6ffd8432.ef6691ea

Iivari, J. (1991). A paradigmatic analysis of contemporary schools of IS development. *European Journal of Information Systems*, *1*(4), 249–272. https://doi.org/10.1057/ejis.1991.47

Iivari, J. (2007). A Paradigmatic Analysis of Information Systems As a Design Science. *Scandinavian Journal of Information Systems*, *19*(2). http://aisel.aisnet.org/sjis/vol19/iss2/5

Iivari, J. (2015). Distinguishing and contrasting two strategies for design science research. *European Journal of Information Systems*, *24*(1), 107–115. https://doi.org/10.1057/ejis.2013.35

Jahn, T., Bergmann, M., & Keil, F. (2012). Transdisciplinarity: Between mainstreaming and marginalization. *Ecological Economics*, *79*, 1–10. https://doi.org/10.1016/j.ecolecon.2012.04.017

James, W. (1922). *Pragmatism: A New Name for Some Old Ways of Thinking*. Dover Publications.

Jirotka, M., Grimpe, B., Stahl, B. C., Eden, G., & Hartswood, M. (2017). Responsible research and innovation in the digital age. *Communications of the ACM*, *60*(5), 62–68. https://doi.org/10.1145/3064940

Jonas, H. (1984). *The imperative of responsibility: in search of an ethics for the technological age*. University of Chicago Press.

Klein, G., Snowden, D., & Pin, C. L. (2010). Anticipatory Thinking. *Informed by knowledge*. Psychology Press.

Kroeze, J. H., Travica, B., & Van Zyl, I. (2019). Information Systems in a Transdisciplinary Perspective: Leaping to a Larger Stage. *Alternation: Interdisciplinary Journal for the Study of the Arts and Humanities in Southern Africa*, *Sp24*. https://doi.org/10.29086/2519-5476/2019/sp24.2a2

Kuechler, W., Vaishnavi, V. K., & Petter, S. (2005). The Aggregate General Design Cycle as a Perspective on the Evolution of Computing Communities of Interest. *Computing Letters*, *1*(3), 123–128. https://doi.org/10.1163/1574040054861221

Kuechler, W., & Vaishnavi, V. (2008). The emergence of design research in information systems in North America. *Journal of Design Research*, *7*(1), 1–16.

Kuhn, T. S. (2012, April 18). *The Structure of Scientific Revolutions: 50th Anniversary Edition*. University of Chicago Press.

Lakatos, I. (1980). *The methodology of scientific research programmes: Philosophical papers* (Vol. 1). Cambridge University Press.

Lee, A. S., Thomas, M., & Baskerville, R. L. (2015). Going back to basics in design science: From the information technology artifact to the information systems artifact. *Information Systems Journal*, *25*(1), 5–21. https://doi.org/10.1111/isj.12054

Levy, M., & Hirschheim, R. (2012). Removing the Positivist Straight Jacket From Information Systems Design Science Research. *ECIS 2012 Proceedings*.

Loorbach, D., Frantzeskaki, N., & Avelino, F. (2017). Sustainability Transitions Research: Transforming Science and Practice for Societal Change. *Annual Review of Environment and Resources*, *42*(1), 599–626. https://doi.org/10.1146/annurev-environ-102014-021340

Lusch, R. F., & Vargo, S. L. (2014a, January 30). *Service-dominant Logic: Premises, Perspectives, Possibilities*. Cambridge University Press.

Lusch, R. F., & Vargo, S. L. (2014b, January 30). *Service-Dominant Logic: Premises, Perspectives, Possibilities*. Cambridge University Press.

Mahfoud, T. (2021). Visions of unification and integration: Building brains and communities in the European Human Brain Project. *New Media & Society*, *23*(2), 322–343. https://doi.org/10.1177/1461444820929576

Majchrzak, A., Markus, M. L., & Wareham, J. (2016). Designing for Digital Transformation: Lessons for Information Systems Research from the Study of ICT and Societal Challenges. *MIS Quarterly*, *40*(2), 267–277. https://doi.org/10.25300/MISQ/2016/40:2.03

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*, 251–266.

McKay, J., Marshall, P., & Hirschheim, R. (2012). The design construct in information systems design science. *Journal of Information Technology*, *27*(2), 125–139.

Mendling, J., Berente, N., Seidel, S., & Grisold, T. (2021). The Philosopher's Corner: Pluralism and Pragmatism in the Information Systems Field: The Case of Research on Business Processes and Organizational Routine. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *52*(2), 127–140. https://doi.org/10.1145/3462766.3462773

Morana, S., vom Brocke, J., Maedche, A., Seidel, S., Adam, M. T. P., Bub, U., Fettke, P., Gau, M., Herwix, A., Mullarkey, M. T., Nguyen, H. D., Sjöström, J., Toreini, P., Wessel, L., & Winter, R. (2018). Tool Support for Design Science Research—Towards a Software Ecosystem: A Report from a DESRIST 2017 Workshop. *Communications of the Association for Information Systems*, *43*(1), 237–256. https://doi.org/10.17705/1CAIS.04317

Morgan, D. L. (2007). Paradigms Lost and Pragmatism Regained. *Journal of Mixed Methods Research*, *1*(1), 48–76.

Morgan, D. L. (2014). Pragmatism as a Paradigm for Social Research. *Qualitative Inquiry*, *20*(8), 1045–1053. https://doi.org/10.1177/1077800413513733

Muehlhauser, L., & Salamon, A. (2012). Intelligence Explosion: Evidence and Import. In A. H. Eden, J. H. Moor, J. H. Søraker, & E. Steinhart (Eds.), *Singularity hypotheses: A scientific and philosophical assessment* (pp. 15–42). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-32560-1_2

Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science research in information systems. *Information & Management*, *51*(6), 801–809. https://doi.org/10.1016/j.im.2014.01.002

Nunamaker, J. F., Briggs, R. O., Derrick, D. C., & Schwabe, G. (2015). The Last Research Mile: Achieving Both Rigor and Relevance in Information Systems Research. *Journal of Management Information Systems*, *32*(3), 10–47.

Nunamaker, J. F., Chen, M., & Purdin, T. D. (1990). Systems Development in Information Systems Research. *Journal of Management Information Systems*, *7*(3), 89–106.

Nunamaker, J. F., Twyman, N. W., & Giboney, J. S. (2013). Breaking out of the Design Science Box: High-Value Impact Through Multidisciplinary Design Science Programs of Research. *AMCIS 2013 Proceedings*, 11.

Nunamaker, J. F., Twyman, N. W., Giboney, J. S., & Briggs, R. O. (2017). Creating High-Value Real-World Impact Through Systematic Programs of Research. *MIS Quarterly*, *41*(2), 335–351.

Nussbaum, M. C. (2023). *Justice for animals: Our collective responsibility*. Simon & Schuster.

O'neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

OpenAI. (2023, March 27). Gpt-4 Technical Report. https://doi.org/10.48550/arXiv.2303.08774

Ord, T. (2020). *The precipice: Existential risk and the future of humanity*. Hachette Books.

Orlikowski, W. J. (2002). Knowing in Practice: Enacting a Collective Capability in Distributed Organizing. *Organization Science*, *13*(3), 249–273. https://doi.org/10.1287/orsc.13.3.249.2776

Orlikowski, W. J., & Iacono, C. S. (2001). Research Commentary: Desperately Seeking the "IT" in IT Research—A Call to Theorizing the IT Artifact. *Information Systems Research*, *12*(2), 121–134. https://doi.org/10.1287/isre.12.2.121.9700

Osterwalder, A., & Pigneur, Y. (2010, July 13). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. John Wiley & Sons.

Ostrom, E. (2010). Beyond Markets and States: Polycentric Governance of Complex Economic Systems. *American Economic Review*, *100*(3), 641–672. https://doi.org/10.1257/aer.100.3.641

Ostrom, E. (2015). *Governing the Commons - The evolution of institutions for collective action*. Cambridge University Press.

Owen, R. (2019). Responsible Innovation and Responsible Research and Innovation. *Handbook on science and public policy* (pp. 26–48). Edward Elgar Publishing. https://doi.org/10.4337/9781784715946.00010

Owen, R., Bessant, J., & Heintz, M. (Eds.). (2013). *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society* (1st ed.). Wiley.

Owen, R., Pansera, M., Macnaghten, P., & Randles, S. (2021). Organisational institutionalisation of responsible innovation. *Research Policy*, *50*(1), 104132. https://doi.org/10.1016/j.respol.2020.104132

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013, April 2). A Framework for Responsible Innovation. In R. Owen, J. Bessant, &

M. Heintz (Eds.), *Responsible innovation: Managing the responsible emergence of science and innovation in society* (pp. 27–50). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118551424.ch2

Paradice, D., Parrish, J., & Richardson, S. (2019). Grand Challenge Pursuits: Insights from a Multi-year DSR Project Stream. *Communications of the Association for Information Systems*, *45*.

Pardales, M. J., & Girod, M. (2006). Community of Inquiry: Its past and present future. *Educational Philosophy and Theory*, *38*(3), 299–309. https://doi.org/10.1111/j.1469-5812.2006.00196.x

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008

Peffers, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: Introduction to the special issue on exemplars and criteria for applicable design science research. *European Journal of Information Systems*, *27*(2), 129–139. https://doi.org/10.1080/0960085x.2018.1458066

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77.

Pennington, D. (2008). Cross-disciplinary Collaboration and Learning. *Ecology and Society*, *13*(2). https://doi.org/10.5751/ES-02520-130208

Pentland, B. T., & Feldman, M. S. (2008). Designing routines: On the folly of designing artifacts, while hoping for patterns of action. *Information and Organization*, *18*(4), 235–250. https://doi.org/10.1016/j.infoandorg.2008.08.001

Popa, F., Guillermin, M., & Dedeurwaerdere, T. (2015). A pragmatist approach to transdisciplinarity in sustainability research: From complex systems theory to reflexive science. *Futures*, *65*, 45–56. https://doi.org/10.1016/j.futures.2014.02.002

Popper, K. (2005). *The Logic of Scientific Discovery*. Routledge.

Porra, J. (2001). A Dialogue with C. West Churchman. *Information Systems Frontiers*, *3*(1), 19–27.

Purao, S., Baldwin, C., Hevner, A. R., Storey, V. C., Pries-Heje, J., Smith, B., & Zhu, Y. (2008). The Sciences of Design: Observations on an Emerging Field. *Communications of the Association for Information Systems*, *23*, 523–546.

Rai, A. (2017a). Diversity of Design Science Research. *MIS Quarterly*, *41*(1), iii–xviii.

Rai, A. (2017b). Editor's Comments: Avoiding Type III Errors: Formulating IS Research Problems that Matter. *MIS Quarterly*, *41*(2), iii–vii.

Rai, A. (2019). Editor's Comments: Engaged Scholarship: Research with Practice for Impact. *MIS Quarterly*, *43*(2), iii–viii.

Ram, S., & Goes, P. B. (2021). Focusing on Programmatic High Impact Information Systems Research, Not Theory, to Address Grand Challenges. *MIS Quarterly*, *45*(1), 479–483.

Reynolds, M., & Holwell, S. (Eds.). (2010). *Systems Approaches to Managing Change: A Practical Guide*. Springer London. Retrieved February 14, 2021, from http://link.springer.com/10.1007/978-1-84882-809-4

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, *4*(2), 155–169.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Schlaile, M. P., Herwix, A., Bogner, K., & Atkins, P. W. B. (2023). An Evolutionary Perspective on Corporate Sustainability Transitions: A Prosocial Approach. In

D. MacKie (Ed.), *The handbook of climate change leadership in organisations: Leadership development in the age of sustainability*. Routledge.

Schlaile, M. P., & Urmetzer, S. (2019). Transitions to Sustainable Development. In W. Leal Filho, A. M. Azul, L. Brandli, P. G. Özuyar, & T. Wall (Eds.), *Decent work and economic growth* (pp. 1–16). Springer International Publishing. https://doi.org/10.1007/978-3-319-71058-7_52-1

Schön, D. A. (1984). *The reflective practitioner: How professionals think in action* (1st ed.). Basic Books.

Schons, M., Pilgram, L., Reese, J.-P., Stecher, M., Anton, G., Appel, K. S., Bahmer, T., Bartschke, A., Bellinghausen, C., Bernemann, I., Brechtel, M., Brinkmann, F., Brünn, C., Dhillon, C., Fiessler, C., Geisler, R., Hamelmann, E., Hansch, S., Hanses, F., ... Group, N. R. (2022). The German National Pandemic Cohort Network (NAPKON): rationale, study design and baseline characteristics. *European Journal of Epidemiology*, *37*(8), 849–870. https://doi.org/10.1007/s10654-022-00896-z

Seger, E., Avin, S., Pearson, G., Briers, M., Ó Heigeartaigh, S., & Bacon, H. (2020). *Tackling threats to informed decision-making in democratic societies – Promoting epistemic security in a technologically-advanced world*. Alan Turing Institute. Retrieved November 4, 2020, from https://www.turing.ac.uk/sites/default/files/2020-10/epistemic-security-report_final.pdf

Seidel, S., Müller-Wienbergen, F., & Rosemann, M. (2010). Pockets of Creativity in Business Processes. *Communications of the Association for Information Systems*, *27*. https://doi.org/10.17705/1CAIS.02723

Seidel, S., & Watson, R. (2020). Integrating Explanatory/Predictive and Prescriptive Science in Information Systems Research. *Communications of the Association for Information Systems*, *47*(1). https://doi.org/10.17705/1CAIS.04714

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, *35*, 37–56.

Serra Undurraga, J. K. A. (2021). What if reflexivity and diffraction intra-act? *International Journal of Qualitative Studies in Education*, *0*(0), 1–14. https://doi.org/10.1080/09518398.2021.1900622

Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization*, *1*(1), 161–176
ZSCC: 0003156.

Simon, H. A. (1973). The Structure of Ill Structured Problems. *Artificial Intelligence*, 21.

Simon, H. A. (1996). *The Sciences of the Artificial* (3rd). MIT Press.

Skaburskis, A. (2008). The Origin of "Wicked Problems". *Planning Theory & Practice*, *9*(2), 277–280. https://doi.org/10.1080/14649350802041654

Stahl, B. C. (2022). Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. *International Journal of Information Management*, *62*, 102441. https://doi.org/10.1016/j.ijinfomgt.2021.102441

Star, S. L. (1989, January 1). Chapter 2 - The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In L. Gasser & M. N. Huhns (Eds.), *Distributed artificial intelligence* (pp. 37–54). Morgan Kaufmann. https://doi.org/10.1016/B978-1-55860-092-8.50006-X

Star, S. L. (2010). This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values*, *35*(5), 601–617. https://doi.org/10.1177/0162243910377624

Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology,

1907-39. *Social Studies of Science*, *19*(3), 387–420. https://doi.org/10.1177/030631289019003001
Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008
Stirling, A. (2006). Precaution, Foresight and Sustainability: Reflection and Reflexivity in the Governance of Science and Technology. *Reflexive governance for sustainable development* (p. 3982). Edward Elgar Publishing. https://doi.org/10.4337/9781847200266.00020
Sun, R., & Gregor, S. (2023). Reconceptualizing platforms in information systems research through the lens of service-dominant logic. *The Journal of Strategic Information Systems*, *32*(3), 101791. https://doi.org/10.1016/j.jsis.2023.101791
Sydelko, P., Midgley, G., & Espinosa, A. (2021). Designing interagency responses to wicked problems: Creating a common, cross-agency understanding. *European Journal of Operational Research*, *294*(1), 250–263. https://doi.org/10.1016/j.ejor.2020.11.045
Tarafdar, M., Shan, G., Bennett Thatcher, J., & Gupta, A. (2022). Intellectual Diversity in IS Research: Discipline-Based Conceptualization and an Illustration from Information Systems Research. *Information Systems Research*, *33*(4), 1490–1510. https://doi.org/10.1287/isre.2022.1176
Tatar, D. (2007). The Design Tensions Framework. *Human–Computer Interaction*, *22*(4), 413–451. https://doi.org/10.1080/07370020701638814
Teo, T. S. H., & Srivastava, S. C. (2007). Information Systems (IS) Discipline Identity: A Review and Framework. *Communications of the Association for Information Systems*, *20*, 29.
Timmermans, J., & Blok, V. (2021). A critical hermeneutic reflection on the paradigm-level assumptions underlying responsible innovation. *Synthese*, *198*(19), 4635–4666. https://doi.org/10.1007/s11229-018-1839-z
Timmermans, J., Blok, V., Braun, R., Wesselink, R., & Nielsen, R. Ø. (2020). Social labs as an inclusive methodology to implement and study social change: The case of responsible research and innovation. *Journal of Responsible Innovation*, *7*(3), 410–426. https://doi.org/10.1080/23299460.2020.1787751
Tremblay, M. C., VanderMeer, D., & Beck, R. (2018). The Effects of the Quantification of Faculty Productivity: Perspectives from the Design Science Research Community. *Communications of the Association for Information Systems*, *43*(1). https://doi.org/10.17705/1CAIS.04334
Turnhout, E. (2009). The effectiveness of boundary objects: The case of ecological indicators. *Science and Public Policy*, *36*(5), 403–412. https://doi.org/10.3152/030234209X442007
Ulrich, W. (2003). Beyond methodology choice: Critical systems thinking as critically systemic discourse. *Journal of the Operational Research Society*, *54*(4), 325–342. https://doi.org/10.1057/palgrave.jors.2601518
Ulrich, W. (2004). Obituary: C West Churchman, 1913–2004. *Journal of the Operational Research Society*, *55*(11), 1123–1129. https://doi.org/10.1057/palgrave.jors.2601825
Ulrich, W. (1994). Can We Secure Future-Responsive Management Through Systems Thinking and Design? *Interfaces*, *24*(4), 26–37. https://doi.org/10.1287/inte.24.4.26
Ulrich, W. (2000). Reflective Practice in the Civil Society: The contribution of critically systemic thinking. *Reflective Practice*, *1*(2), 247–268. https://doi.org/10.1080/713693151

Ulrich, W. (2001). Critically Systemic Discourse: A Discursive Approach to Reflective Practice in ISD (Part 2). *The Journal of Information Technology Theory and Application (JITTA)*, *3*(3), 85–106.

Ulrich, W. (2006). Critical Pragmatism: A New Approach to Professional and Business Ethics. *Interdisciplinary yearbook for business ethics. v. 1, v. 1.* Peter Lang Pub Inc.

Ulrich, W. (2012a). Operational research and critical systems thinking – an integrated perspective. Part 1: OR as applied systems thinking. *Journal of the Operational Research Society*, *63*(9), 1228–1247. https://doi.org/10.1057/jors.2011.141

Ulrich, W. (2012b). Operational research and critical systems thinking—an integrated perspective. Part 2: OR as argumentative practice. *Journal of the Operational Research Society*, *63*(9), 1307–1322. https://doi.org/10.1057/jors.2011.145

Ulrich, W. (2016). Philosophy for professionals: Towards critical pragmatism [magazine]. *Ulrich's Bimonthly*, *14*(2), 1–17.

Ulrich, W., & Reynolds, M. (2010). Critical Systems Heuristics. In M. Reynolds & S. Holwell (Eds.), *Systems approaches to managing change: A practical guide* (pp. 243–292). Springer London. https://doi.org/10.1007/978-1-84882-809-4_6

UNESCO. (2017). Annex II: Recommendation on Science and Scientific Researchers.

Vaishnavi, V., Kuechler, B., & Petter, S. (2019). *Design Science Research in Information Systems*. Retrieved October 28, 2020, from http://www.desrist.org/desrist/content/design-science-research-in-information-systems.pdf

Vaishnavi, V., & Kuechler, W. J. (2015). *Design Science Research Methods and Patterns* (2nd Edition). CRC Press.

Van de Ven, A. H. (2018). Academic-practitioner engaged scholarship. *Information and Organization*, *28*(1), 37–43. https://doi.org/10.1016/j.infoandorg.2018.02.002

van Gigch, J. P., Koenigsberg, E., Dean, B., & Churchman, C. W. (1997). In Search of an Ethical Science: An Interview with C. West Churchman an 80th Birthday Celebration. *Journal of Business Ethics*, *16*(7), 731–744.

Vargo, S. L., Koskela-Huotari, K., & Vink, J. (2020). Service-dominant logic: Foundations and applications. *The Routledge Handbook of Service Research Insights and Ideas*. Routledge.

Vargo, S. L., & Lusch, R. F. (2016). Institutions and axioms: An extension and update of service-dominant logic. *Journal of the Academy of Marketing Science*, *44*(1), 5–23. https://doi.org/10.1007/s11747-015-0456-3

Venable, J., Pries-Heje, J., & Baskerville, R. (2014). Feds: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Vervaeke, J., Lillicrap, T. P., & Richards, B. A. (2012). Relevance Realization and the Emerging Framework in Cognitive Science. *Journal of Logic and Computation*, *22*(1), 79–99. https://doi.org/10.1093/logcom/exp067

Vervaeke, J., & Ferraro, L. (2013a). Relevance Realization and the Neurodynamics and Neuroconnectivity of General Intelligence. In I. Harvey, A. Cavoukian, G. Tomko, D. Borrett, H. Kwan, & D. Hatzinakos (Eds.), *Smartdata* (pp. 57–68). Springer New York. https://doi.org/10.1007/978-1-4614-6409-9_6

Vervaeke, J., & Ferraro, L. (2013b). Relevance, Meaning and the Cognitive Science of Wisdom. In M. Ferrari & N. M. Weststrate (Eds.), *The scientific study of personal wisdom* (pp. 21–51). Springer Netherlands. https://doi.org/10.1007/978-94-007-7987-7_2

Vervaeke, J., & Mastropietro, C. (2021). Gnosis in the Second Person. *Dispatches from a time between worlds: Crisis and emergence in metamodernity*. Perspectiva.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*(1), 233. https://doi.org/10.1038/s41467-019-14108-y

vom Brocke, J., Stein, A., Hofmann, S., & Tumbas, S. (2015). *Grand Societal Challenges in Information Systems Research and Education: Ideas from the ERCIS Virtual Seminar Series*. Springer International Publishing. Retrieved December 13, 2018, from //www.springer.com/gp/book/9783319150260

vom Brocke, J., Weber, M., & Grisold, T. (2021). Design Science Research of High Practical Relevance. In S. Aier, P. Rohner, & J. Schelp (Eds.), *Engineering the Transformation of the Enterprise: A Design Science Research Perspective* (pp. 115–135). Springer International Publishing. https://doi.org/10.1007/978-3-030-84655-8_8

vom Brocke, J., Winter, R., Hevner, A. R., & Maedche, A. (2020). Accumulation and Evolution of Design Knowledge in Design Science Research – A Journey Through Time and Space. *Journal of the Association for Information Systems*, *21*(3), 520–544. https://doi.org/10.17705/1jais.00611

von Schomberg, R. (2013, May 20). A Vision of Responsible Research and Innovation. *Responsible innovation: Managing the responsible emergence of science and innovation in society*. Wiley. Retrieved January 6, 2020, from https://papers.ssrn.com/abstract=2428157

von Schomberg, R. (2019). Why responsible innovation? In R. von Schomberg & J. Hankins (Eds.), *International handbook on responsible innovation: A global resource* (pp. 12–32). Edward Elgar Publishing Cheltenham, UK.

von Schomberg, R., & Hankins, J. (2019, July 26). *International Handbook on Responsible Innovation. A Global Resource*. https://doi.org/10.4337/9781784718862

Walsham, G. (2017). Ict4d research: Reflections on history and future agenda. *Information Technology for Development*, *23*(1), 18–41. https://doi.org/10.1080/02681102.2016.1246406

Wang, P. (2021). Connecting the Parts with the Whole: Toward an Information Ecology Theory of Digital Innovation Ecosystems. *MIS Quarterly*, *45*(1), 397–422.

Wanzenböck, I., Wesseling, J. H., Frenken, K., Hekkert, M. P., & Weber, K. M. (2020). A framework for mission-oriented innovation policy: Alternative pathways through the problem–solution space. *Science and Public Policy*, scaa027. https://doi.org/10.1093/scipol/scaa027

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science*, *16*(4), 409–421. https://doi.org/10.1287/orsc.1050.0133

Wieringa, R. (2009). Design science as nested problem solving. *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology - DESRIST '09*. https://doi.org/10.1145/1555619.1555630

Wilson, D. S., Madhavan, G., Gelfand, M. J., Hayes, S. C., Atkins, P. W. B., & Colwell, R. R. (2023). Multilevel cultural evolution: From new theory to practical applications. *Proceedings of the National Academy of Sciences*, *120*(16), e2218222120. https://doi.org/10.1073/pnas.2218222120

Wilson, D. S., Ostrom, E., & Cox, M. E. (2013). Generalizing the core design principles for the efficacy of groups. *Journal of Economic Behavior & Organization*, *90*, S21–S32. https://doi.org/10.1016/j.jebo.2012.12.010

Wilson, D. S., Philip, M. M., MacDonald, I. F., Atkins, P. W. B., & Kniffin, K. M. (2020). Core design principles for nurturing organization-level selection. *Scientific Reports*, *10*(1), 1–6. https://doi.org/10.1038/s41598-020-70632-8

Wilson, E. O. (1999, March 30). *Consilience: The Unity of Knowledge* (Reprint edition). Vintage.

Winter, S. J., & Butler, B. S. (2011). Creating Bigger Problems: Grand Challenges as Boundary Objects and the Legitimacy of the Information Systems Field. *Journal of Information Technology*, *26*(2), 99–108. https://doi.org/10.1057/jit.2011.6

Wittmayer, J. M., & Schäpke, N. (2014). Action, research and participation: Roles of researchers in sustainability transitions. *Sustainability Science*, *9*(4), 483–496. https://doi.org/10.1007/s11625-014-0258-4

Yudkowsky, E. (2023). The Open Letter on AI Doesn't Go Far Enough [magazine]. *Time*. Retrieved April 8, 2023, from https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/

Zuboff, S. (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

What Happens When the Machines Stop? Uncovering the Risk of Digital Fragility as the Achilles' Heel of the Digital Transformation of Societies

Outline

Abstract

The digital transformation of societies has been a core concern for the information systems (IS) research community since its emergence. While most of this work has had a positive outlook, recently a stronger focus on the unintended consequences and dark side of digitalization has come to the fore. This paper contributes to this emerging stream of research by zooming in on a heretofore unrecognized question with potentially catastrophic consequences: What happens to our increasingly digitalized societies when a prolonged blackout causes a large fraction of digital systems and services to stop working for an extended period of time? To answer this motivating question, we conducted two system dynamics-based simulation experiments to tease out how different degrees of digitalization in a society would affect the resilience of the food system in the face of two different, extreme but plausible prolonged blackout scenarios. We find that a high degree of digitalization has a strong significant negative impact on food system resilience in the investigated scenarios. In the discussion of our findings, we conceptualize "the risk of digital fragility" as the underlying driver of the observed results. Moving forward, we suggest seven mitigation strategies for the risk of digital fragility as fruitful avenues for future research.

Keywords: Digital transformation, digital agility, digital fragility, systemic catastrophic risk, blackout, system dynamics, simulation experiment.

Bibliographic Information

Herwix, A., Tieman, R., Rivers, M., Rosenkranz, C., & Denkenberger, D. (2023). What Happens When the Machines Stop? Uncovering the Risk of Digital Fragility as the Achilles' Heel of the Digital Transformation of Societies. *SocArXiv*.
https://doi.org/10.31235/osf.io/2vctm

Author's contribution

Conceptualization (lead), Methodology (lead), Software (lead), Validation (equal), Formal analysis (lead), Investigation (equal), Data curation (lead), Writing – original draft (lead), Writing – review and editing (lead), Visualization (lead), Project administration (equal).

1 Introduction

> "The best way to keep something bad from happening is to see it ahead of time... and you can't see it if you refuse to face the possibility."
>
> — William S. Burroughs

The study of the development, adoption, use and impact of digital systems and services through a socio-technical lens has been the defining feature of the IS research community at large (Sarker et al., 2019). While most of this work is motivated by and grounded in the *positive* expectation that digital systems and services can help to improve personal, organizational or societal agility and performance, our research community is also increasingly becoming aware that the pervasive use of digital systems and services may also have potential *negative* effects, often referred to as "unintended consequences" or "dark side" (e.g., D'Arcy et al., 2012; Giermindl et al., 2022; Mikalef et al., 2022; Tarafdar et al., 2013). This includes, for example, research on technostress (e.g., Tarafdar et al., 2019), the potential dark side of social media on democratic decision making (e.g., Seger et al., 2020), or the misuse of artificial intelligence (AI) technologies (e.g., Mikalef et al., 2022). Importantly, such research has helped us gain a deeper and more nuanced understanding of the costs and risks that ought to be considered as the digital transformation of our societies[1] continues. We aim to contribute to this valuable stream of research and argue that IS research so far has overlooked at least one fundamental question with potentially catastrophic consequences: What happens to our increasingly digitalized societies when a prolonged blackout causes a large fraction of digital systems and services to stop working for an extended period of time?

Digital systems and services are inherently dependent on electricity, which makes them generally strongly vulnerable to blackouts given the centralized design of most societies electricity infrastructures. Without access to the electricity grid, it is estimated that digital systems and services will start failing within minutes. Most digital systems and services would be reaching complete failure within just a few days (i.e., after backup power generators run out of fuel or start failing themselves; Petermann et al., 2011; Stockton & Council, 2016, 2018). Thus, we argue that it is critical and prudent to consider the potential negative effects that pervasive use and dependency on digital systems and services may have in prolonged blackout scenarios.

In particular, what if the digital transformation of our societies is not only making our organizations more agile, productive, and resilient but also has the unintended dark side of making our societies more *fragile* and *vulnerable* in catastrophic scenarios involving blackouts? We deem this to be an important avenue to investigate as we continue to rapidly increase the dependencies of our societies on digital systems and services. In our view, as IS researchers, we have a moral responsibility as well as the necessary expertise to engage with such concerns and help our societies to consciously manage the tensions between the potential bright sides and dark sides of the digital transformation.

As a first step towards engaging with this line of thinking, we seek to rigorously validate and test our premises. For this, we employ a devil's advocate perspective and explore the impact of the digital transformation in extreme but plausible blackout scenarios in which a large fraction of digital systems and services would stop working. We refer to these

[1] Throughout this paper we use the phrases *digital transformation of societies* and *digital transformation* interchangeably to refer to the overarching societal transformation induced by the rapid and ongoing digitalization of processes, digitization of data, and digital transformation of organizations (Hanelt et al., 2021; Vial, 2019; Wessel et al., 2021).

prolonged blackouts as catastrophic electricity loss (CEL) scenarios[2] that may be triggered by events such as a coordinated cyberattack or a series of high-altitude electromagnetic pulses (HEMPs) caused by the use of nuclear weapons. While it might be uncomfortable to consider such dark scenarios and tempting to dismiss them as highly unlikely, given their outsized potential impact[3] and the limited attention we tend to afford them, it is still considered to be of high expected value to engage with them (Dolan, 2018; Petermann et al., 2011; Stockton & Council, 2016). We do this by using a system dynamics (SD)-based Monte Carlo simulation of the U.S. food supply chain to investigate the research question: *How do different degrees of digitalization in a society affect the resilience of the U.S. food system in catastrophic electricity loss scenarios?* This is an appropriate research approach to engage with our overarching concern as it allows us to rigorously explicate, quantify, and interrogate our assumptions regarding the potential negative impact of the digital transformation of societies in CEL scenarios in a domain of utmost societal relevance. In particular, the food system is generally designated as one of the critical infrastructures (CIs) of society that "provide goods and services that enable the maintenance and sustainment of public wellbeing including public safety, economic vitality, and security" (Katina & Keating, 2015, p. 317).

We find that in all of our simulated scenarios, a higher degree of digitalization indeed leads to significant worse outcomes in terms of the peak amount of people affected by food shortages as well as the absolute number of days where people did not have access to food.

Based on our scenario investigations and simulation results, we then go beyond the particularities of CEL scenarios and develop a novel theoretical framing that makes sense of the general dynamics we have observed. Specifically, we identify *the promise of digital agility* as a driving force and the potential bright side of the digital transformation of societies that is tempered by *the risk of digital fragility* as a potential dark side. As our results suggest, managing the interplay of these two sides will be a key challenge for our societies in general and IS research in particular. To kickstart engagement with this challenge, we aim to guide future research with a first set of tentative strategies that could help to manage it productively.

Our work contributes a fresh and critical look on a potential dark side of the digital transformation of societies that has largely been overlooked by IS research. Specifically, we illustrate the importance of considering CEL scenarios in the context of the ongoing digital transformation of our societies. Moreover, abstracting from the specifics of our work, we identify a set of general underlying dynamics inherent to the digital transformation of our societies that explain our results as a form of systemic catastrophic risk on the macro scale induced by competitive dynamics on the meso scale. Thus, we provide a simple but coherent explanation that resolves the mystery of how digital transformation with its promise of digital agility inadvertently ends up contributing to the risk of digital fragility. The usability and utility of our framing is demonstrated by a set of generic mitigation strategies for the risk of digital fragility that provide fruitful avenues for future research.

The rest of the paper is structured as follows. First, we clarify the key concepts used in this study in the theoretical background section. Then, we explain our research approach. Third, we present our scenario development and the results of two simulation experiments. Fourth, we discuss our findings and delineate implications for IS research, IS practice, and policy making. Fifth, we conclude the paper with a call to action.

[2]With catastrophic electricity loss we mean blackouts covering a large geographic area with the size of multiple US states or EU countries, affecting at least 90% of the population and lasting for at least 30 days.

[3]One just has to look at the recent blackouts in Texas in February 2021 to get a glimpse of the potentially catastrophic consequences that large-scale and prolonged power outages can cause (Traywick et al., 2021).

2 Theoretical Background

2.1 The Digital Transformation of Societies and the Promise of Digital Agility

The digital transformation of our societies is quickly progressing as more and more aspects of our lives are becoming pervaded by digital systems (Haigh, 2022; Yoo, 2010). While this phenomenon may be examined from multiple viewpoints, in this paper, we are mostly concerned with a macro (i.e., societal) level perspective on the digital transformation of societies as a process of systemic and societal change that emerges from the interplay of competitive dynamics on the meso (i.e., organizational) level which, in turn, emerge from actions taken by individual actors on the micro level (Dopfer et al., 2004). Taking the macro level perspective, it has been suggested that much of the momentum behind the digital transformation of our societies is fueled by, what we call, *the promise of digital agility* (i.e., competitive advantage and resilience to be derived from the effective use of digital systems and services) on the meso level (Hanelt et al., 2021; Vial, 2019; Wessel et al., 2021). As illustrated in Figure 16, we view this promise of digital agility as being grounded in the realization that digital systems and services afford standardization and interconnectedness, which may be leveraged for competitive advantage and organizational resilience in competitive organizational environments.

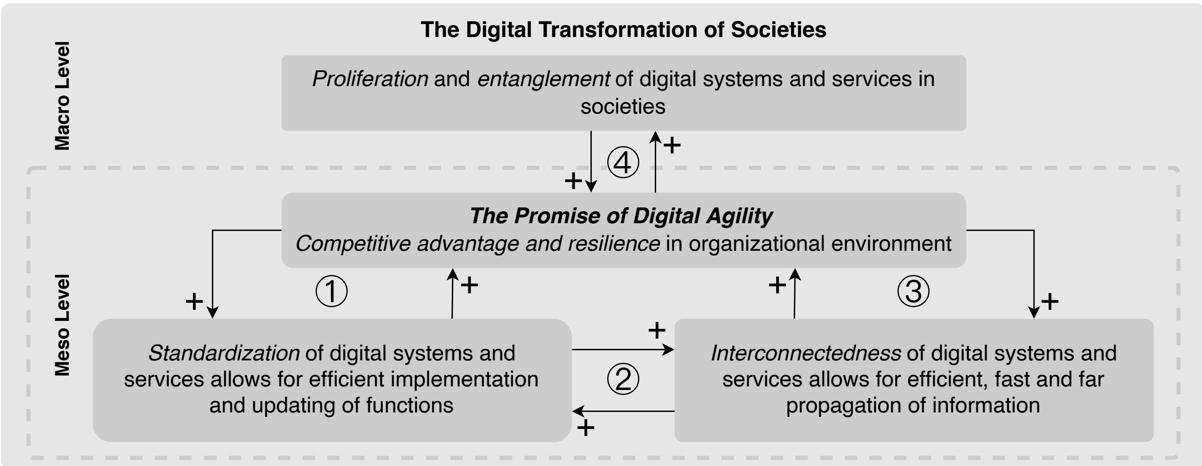

Figure 16: The promise of digital agility is grounded in the standardization and interconnectedness afforded by digital systems and services. Numbered relationships are explained in the text.

Referring to (1) in Figure 16, the *standardization* of modern digital systems and services generally follows a modular layered architecture (Yoo et al., 2010), which allows for the efficient implementation and continued updating of a wide variety of behaviors and functions. For instance, the same computer chip can be used for multiple different purposes, according to the software uploaded on it or a piece of software can be copied and installed on any number of machines. This organizing logic is so useful for business purposes that it is generally acknowledged that it confers competitive advantage to those who are able to leverage it most effectively (Yoo et al., 2010). This realization, in turn, has been sparking further investments into the development and standardization of a broad variety of digital systems and services (Haigh, 2022).

Related to (2) in Figure 16, the standardization of modern digital systems and services allows for and supports and unprecedented degree of *interconnectedness*, which, in turn, allows for the organization and standardization of even more complex and powerful digital systems and services (Sandberg et al., 2020). For instance, in the emerging Internet of Things (IoT) digital systems and services are now starting to be used to network entire cities together to allow for a greater control of processes, increased automation, and remote access (Zanella et al., 2014).

Regarding (3) in Figure 16, the increasing interconnectedness of digital systems and services also has important business value as it allows for a more efficient information transfer. Thus, the effective leveraging of interconnectedness promises competitive advantage, which, in turn, is sparking further investments to develop the interconnectedness of digital systems and services (Sandberg et al., 2020), for instance, as part of the emerging IoT (Zanella et al., 2014).

Considering (4) in Figure 16, all of these features and relationships fuel the promise of digital agility and, thus, contribute to a *proliferation* and *entanglement* of interconnected digital systems and services in society (Hillis, 2010). In a positive reinforcing feedback loop, this proliferation and entanglement, in turn, creates a competitive landscape that further grows the interest in the promise of digital agility as organizations increasingly fear to be left behind.

Importantly, these dynamics of the digital transformation contribute to an increasing set of complex interdependencies between different sectors of societies. For instance, digital systems and services rely on the electricity grid, which in turn depends more and more on digital systems and services for its functioning (Korkali et al., 2017; Parandehgheibi & Modiano, 2013; Siegel, 2018). Additionally, the electricity grid depends on power generation, which requires fossil fuel supply networks or renewable energy production systems, which in turn depend on digital supervisory control and data acquisition (SCADA) systems (Rinaldi, 2004).

As a result, the most critical sectors of our societies are increasingly viewed, modeled, and analyzed collectively as interdependent complex networks or *networks of networks* (Gao et al., 2012; Kenett et al., 2015) rather than mostly independent sectors that are only loosely coupled. This change in perspective is pivotal because it emphasizes the potential for failures that can cascade across many different critical sectors of society, a fact which has already been observed in a variety of real life cases (Chang et al., 2007; Dobson et al., 2007; Haes Alhelou et al., 2019).

Given the potentially catastrophic outcomes of cascading failures (Petermann et al., 2011; Stockton & Council, 2016), scholars have started to call for a more systematic engagement with the mechanisms and drivers in such scenarios. Explicitly called for are the development of useful models of interdependent critical infrastructures (Helbing, 2013) as well as guidelines for their design, regulation, and governance (Centeno et al., 2015; Hollick & Katzenbeisser, 2019). In particular, due to the nature of the challenge, more researcher attention from disciplines with experience in socio-technical research perspectives is needed to prepare for or help prevent the most catastrophic scenarios that cascading failures across critical sectors of societies could entail (Centeno et al., 2015; Dolan, 2018; Helbing, 2013).

2.2 Cascading Failures and the Threat of Catastrophic Electricity Loss

Catastrophic electricity loss (CEL) is a phrase that we have coined for this paper. It refers to prolonged power outages of catastrophic proportions, sometimes also referred to as "black sky" events (e.g., Monken, 2015; Stockton & Council, 2016). As a general rule of thumb, we start to speak of CEL when 90% of customers across a multistate (US-State) area lose electricity access for at least 30 days with long-term demand for emergency power (i.e., power provided by backup generators and alternative sources) and the potential for a slow recovery of grid capacity (sic., up to several months or years) to pre-event levels (NERC, 2012; Stockton & Council, 2016). As such, it has to be observed that despite several large scale (and already devastating) power outages across different parts of the world (Haes Alhelou et al., 2019), an event that would qualify as CEL has never happened up to now. Thus, CEL is, as of now, a truly rare and extreme event, which makes it hard to predict and easy to dismiss as unlikely or unrealistic (Goodwin & Wright, 2010; G. Wright & Goodwin, 2009). However, researchers highlight that CEL is certainly plausible (some say it is even inevitable; Dolan, 2018) and should be prepared for given its potentially catastrophic impact on societies (Hollick & Katzenbeisser, 2019; Petermann et al., 2011; Stockton & Council, 2016).

Previous real world large-scale electricity loss such as during and after hurricane Katrina in 2005 (Reed et al., 2010) or during and after hurricane Maria in 2017 (Román et al., 2019) already illustrate the real harm and hardship that large-scale and long-term power outages can mean for people and societies. However, the impact of a CEL event would be much larger than these already devastating events. CEL could even occur on a global scale (e.g., in nuclear war or solar storm scenarios) and, thus, pose a profound challenge to humanity as a whole because significant shifts in global system dynamics (e.g., due to tipping points being crossed; Gladwell, 2000) would need to be expected. For instance, there are inherent and significant challenges to restarting entire electricity grids (Good, 2012; Siegel, 2018) or even multiple critical sectors of societies from complete shutdowns (Maher & Baum, 2013; Petermann et al., 2011).

Due to digital systems and services ultimate dependency on electric energy, we see the potential for such profound negative consequences as a critical incentive to engage with the threat of CEL scenarios. In particular, we chose to engage with CEL scenarios because they are generally highly neglected but tractable, which indicates a high cost effectiveness given their potential impact (Herwix & Haj-Bolouri, 2021). For instance, previous work has sketched out how food might be provided in the event of global CEL (Cole et al., 2016) and how other needs might be met (Abdelkhaliq et al., 2016). Further work explored scenarios of sun-obscuring catastrophes (e.g., asteroid or comet impact, supervolcanic eruption, and nuclear winter), which could cascade into near global CEL (Denkenberger et al., 2017). However, to the best of our knowledge no prior research has systematically considered the role of the digital transformation of societies in CEL scenarios.[4] This is an important gap given well-established concerns about the increasing interdependence between critical sectors of our societies (e.g., Buldyrev et al., 2010; Korkali et al., 2017; Rinaldi et al., 2001), which is only predicted to deepen as the digital transformation of our societies continues (e.g., Onyeji et al., 2014; Wang & Lu, 2013).

3 Research Approach

The research leading up to this paper can best be described as a combination of *scenario planning* in the *intuitive logic* school (Amer et al., 2013; Wilkinson et al., 2013) with a *system dynamics* modeling and simulation exercise (Fang et al., 2018). This is an appropriate research approach to answer our research question because scenario planning allows us to work with rare and difficult to predict events (Derbyshire & Wright, 2014; Goodwin & Wright, 2010; G. Wright & Goodwin, 2009) whereas SD modelling provides us with a rigorous way to open up our reasoning to critical scrutiny and academic discussion as well as play out the consequences of our assumptions (Fang et al., 2018; Rahmandad &

[4]See Appendix A for the results of a systematic review of IS research on this topic.

Sterman, 2012).

Specifically, we use scenario planning in the intuitive logic school to identify and explore plausible scenarios of CEL—"a set of hypothetical events set in the future constructed to clarify a possible chain of causal events as well as their decision points" (Kahn & Wiener, 1967, p. 6)—with the goal of better understanding how a hypothetical situation might come about and be influenced rather than making quantitative assessments of their likelihood (Amer et al., 2013; Kahn & Wiener, 1967). This is reasonable because while CEL scenarios are inherently complex, rare and uncertain events and, thus, difficult to predict with a well-calibrated probability distribution[5], this does not mean that their occurrence is unlikely (Goodwin & Wright, 2010). Even events which have been rare in the past, can still occur with a very high probability in the near future.[6]

Thus, we view our scenarios not as predictions of the future but rather as fictions and stories that can be deployed to question assumptions, start discussions, and work toward shared interpretative frames (Goodwin & Wright, 2010; Wilkinson et al., 2013; A. Wright, 2005). They should be understood as reframing devices rather than forecasting tools and not be judged in terms of their predictive accuracy but by their utility for action (Wilkinson et al., 2013). In our case, we used the developed scenarios as inputs for the adaptation of a SD-based simulation model of the U.S. food supply chain[7] to assess its resilience in the face of CEL scenarios. By doing so, we are able to combine the flexibility of the scenario planning approach with the rigorousness, extensibility, and auditability of simulation-based research (Dong, 2022).

Process-wise, our research was realized through a multi-step approach visualized in Figure 17. We started with the identification of plausible scenarios that fit our definition of CEL. Here, we followed prior suggestions to employ a *devil's advocate* perspective to consider rare but influential events that could have potentially disastrous effects (Goodwin & Wright, 2010; G. Wright & Goodwin, 2009). This can help to challenge preexisting assumptions about likely futures, which has been argued to improve decision making by opening up horizons, challenging group think and combatting frame blindness (Goodwin & Wright, 2010; G. Wright & Goodwin, 2009). For this we have, in line with best practice recommendations (G. Wright & Goodwin, 2009), developed a scenario framework that establishes a systematic understanding of the progression of catastrophic risks and helps us in developing the broad outlines of scenarios which could cause CEL. In a next step, we then selected two potential scenarios for deeper investigation. Our choice of scenarios can be characterized as a convenience sample influenced by fit with our research question, expertise available, and length restrictions for this paper.

We then engaged in desk research as well as a series of informal workshop sessions to investigate and develop the causal logic within each of the two scenarios. This step was informed and guided by broad but extensive literature searches on the topics as well as prior work of the authors that already engaged with the subject matter (e.g., doing impact analyses of HEMPs and other catastrophes – sources redacted for review). We used a comprehensive network diagram as well as specific causal diagrams as collaborative

[5]A probability distribution is called calibrated if the chance that we assign to an event is accurate. For instance, it snows on 5% of days when we estimated the probability of snow is 5%. If it snows on more, or less, than 5% of those days then the probability assessment is mis-calibrated (Goodwin & Wright, 2010).

[6]For instance, taking the seminal example of the black swan (Taleb, 2010), there were hundreds of years where Europeans had a very low probability of encountering a black swan, simply because they had never visited large parts of the world. But once expeditions to Australia where getting more common, the first black swan encounter was simply a matter of time. However, given the data available at the time, the increasing chance of a black swan encounter would have been almost impossible to predict.

[7]We have adapted a peer-reviewed system dynamics based model of the U.S. food supply chain that was developed to assess the resilience of the U.S. food system in severe pandemics (Huff et al., 2015).

objects (Barley et al., 2012) to integrate our understanding. After several rounds of discussions within the author team as well as consultations with two outside experts with multiple years of professional experience in relevant domains[8], the authors reached consensual agreement regarding the causal logic within the scenarios and decided on a set of key parameter ranges (sic., severity of disruption and minimum recovery time for key affected sectors of society) that would provide plausible boundaries for the two selected CEL scenarios.

Next, we focused on the adaptation of an existing, peer-reviewed SD-based model of the U.S. food supply chain (Huff et al., 2015) to simulate how different degrees of digitalization would affect the outcome of CEL scenarios in a critical sector of society. This phase was highly iterative with simulation model development and validation going hand in hand. We used the recommended model validation guidelines from Fang et al. (2018) to ensure the validity and usefulness of our simulation model as summarized in Appendix B. In particular, we constructed two baseline blackout scenarios that more closely resemble already observed (i.e., less extreme) blackouts so that we could broadly compare the outputs of the model against available data. The details of these behavior tests are summarized in the Online Appendix.

Once all the validation tests passed and we were confident in the usefulness of our simulation model, we designed two experiments that would help us answer our research question. For each selected scenario we estimated a set of parameter ranges that would allow us to simulate their outcomes for two conditions: (1) Our current estimated level of digitalization, and (2) a higher level of digitalization that could plausibly come to be realized within the next decade. Thus, in line with general practice in simulation research, we decided on a simple experimental design with one control and one treatment condition that we could test with statistical hypothesis significance testing.

For the data generation we used the sensitivity simulation facility in the modelling and simulation software VENSIM PLE+[9] to run two reproducible Monte Carlo simulations with 2000 iterations per experiment. Using Monte Carlo simulations with thousands of iterations enabled us to sample a large fraction of the plausible scenario space as it allows for the use of probability distributions over specified ranges to estimate the interplay between uncertain parameters rather than having to settle for less informative point estimates for all simulation parameters. Thus, Monte Carlo simulation is a particularly appropriate method to investigate inherently uncertain phenomena such as CEL scenarios.

For the data analysis, the data generated by VENSIM PLE+ was exported to Microsoft Excel to calculate descriptive statistics as well as to conduct the statistical hypothesis significance testing. In particular, we used the built-in Excel functions for F-tests and two-tailed t-tests to calculate the p-values necessary for the testing of our hypotheses. All of the necessary data to reproduce our analyses including the simulation model, a documentation of the simulation model, a justification for our parameter choices, configuration files for running our experiments, the Monte Carlo simulation results, and the Excel spreadsheet used for data analysis are made accessible as an Online Appendix to this publication.

In a final step of synthesis, we then reflected on the results of our experiments and scenario investigations and identified underlying dynamics in the considered CEL scenarios. We conceptualized these dynamics as contributing to "the risk of digital fragility" which drives and is driven by systemic catastrophic risks for our societies. Tentative mitigation strategies for digital fragility were derived based on this framing as well as related work.

[8]We consulted an academic expert in cybersecurity as well as an professional working at an electric utility in a relevant capacity.

[9]https://vensim.com/

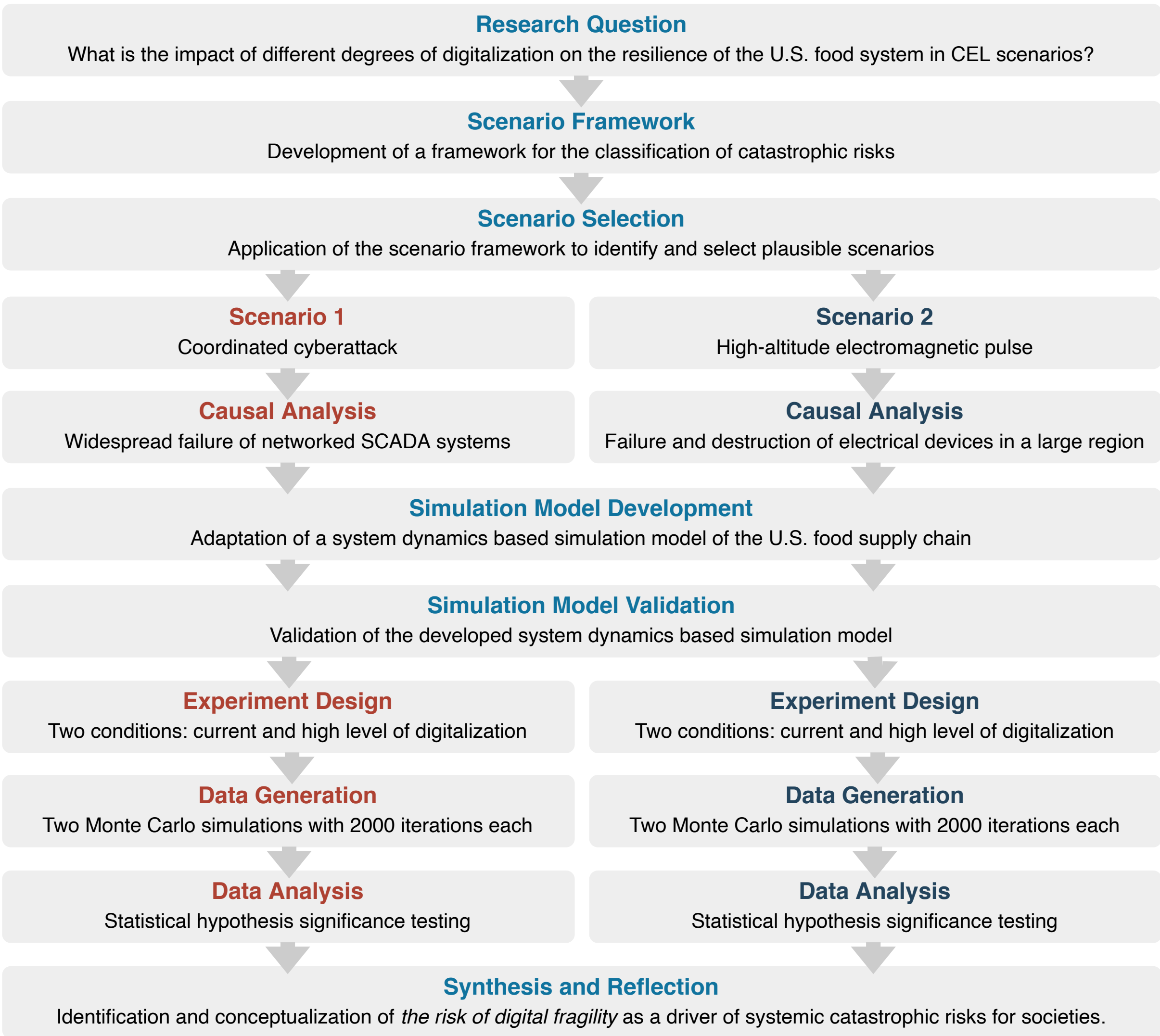

Figure 17: Summary of the Research Process.

4 Scenario Development

Our scenario development was guided by a comprehensive scenario framework, which we present in depth in Appendix C. In short, the framework opened up a potential space of 63 unique broad catastrophic risk scenarios along the three key dimensions of *risk origin* (how does the risk begin?), *risk scaling* (how does the risk reach catastrophic scale?), and *risk impact* (how does the risk impact a focal system?).[10] Given these key scenario dimensions, we were able to systematically develop the broad outlines of a variety of plausible CEL scenarios by considering combinations of characteristics. In particular, we developed several broad classes of plausible CEL scenarios,[11] but decided to focus on two scenario classes that exhibited strong ties to the digital transformation of our societies and, thus, were particularly relevant to our investigation (see Table 12):

- *Coordinated cyberattacks* intentionally caused by a small number or a large number of humans (i.e., a malicious risk or a conflict risk) as a scenario that is deemed to become more likely as cyberwarfare is proliferating (e.g., A. King & Gallagher, 2020),
- *High-altitude Electro-Magnetic Pulse,* either, intentionally caused by humans, whether small in numbers (i.e., a malicious risk) or large in numbers (i.e., a conflict risk), or unintentionally caused by them (i.e., an accident risk) as a scenario where digital systems may be rapidly destroyed on a catastrophic scale (e.g., Wilson, 2008).

Table 12: An overview of the investigated scenarios with a classification of their associated risk types with the electricity grid as the focal system.

Scenario Name	**Risk Origin**	**Risk Scaling**	**Risk Impact**
Coordinated Cyberattacks	Malicious Risk, Conflict Risk	Cascading Risk	Functioning Risk, Infrastructure Risk
High-altitude Electro-Magnetic Pulse	Malicious Risk, Conflict Risk, Accident Risk	Leverage Risk	Functioning Risk, Infrastructure Risk

4.1 Coordinated Cyberattacks

For our simulation experiment, we consider a coordinated cyberattack on the electrical grid that was intentionally caused by a small number of humans (e.g., an elite cyberwarfare team of an adversarial state) as this seems to be the most plausible way in which a

[10] In particular, the risk origin dimension highlights seven potential starting points for a CEL scenario. The risk scaling dimension highlights three potential ways a CEL scenario could reach a catastrophic scale. The risk impact dimension highlights three potential ways a focal system could be catastrophically impacted.

[11] We considered amongst others: a severe pandemic caused without human involvement (i.e., a natural risk) or caused by humans, whether intentionally (i.e., a malicious risk) or unintentionally (i.e., an accident risk); a hurricane as a naturally occurring phenomenon (i.e., a natural risk) that could be intensified or even caused by human involvement in the climate and weather system due to massive greenhouse gas emissions (i.e., a commons risk), a geomagnetic storm as a naturally occurring phenomenon (i.e., a natural risk).

cyberattack could cause CEL.[12] The U.S. Electric Infrastructure Security (EIS) council notes that with the rise of terrorism and the sophistication of cyberwarfare measures, it is prudent to assume that eventually a large scale cyberattack will cause a wide-scale power outage (Stockton & Council, 2016). In particular, a coordinated attack to disable the supervisory control and data acquisition (SCADA) systems (i.e., systems that control the operation of machinery) of power plants or in other electrical grid components should be treated as a plausible scenario given that it could be perpetrated not only by nation states but also cyber terrorist groups. Attacks by cyber terrorist groups are very difficult to deter because retaliation is hampered by a lack of international treaties and enforcement mechanisms against such groups (Tehrani et al., 2013). Importantly, such attacks need not necessarily be limited to a given country or region as SCADA systems are increasingly interconnected through proprietary networks (Korkali et al., 2017) and sometimes even reachable via the Internet (Pliatsios et al., 2020). In our hypothetical scenario a 'Stuxnet' like worm (R. King, 2012; Nicolas et al., 2011) infiltrates a large fraction of SCADA systems involved in the electricity grid and then triggers a simultaneous disruption that almost instantaneously takes down around 90% of the entire U.S. electricity grid. Attempts to recover the electricity grid start immediately but the sheer size of the outage and failures in backup power generators at some facilities lead to severe challenges in communication, coordination and recovery (Siegel, 2018). This is troublesome because recovery from complete shutdown of the electricity grid is a complex process that requires intensive communication to sync up different sections of the electricity grid (Good, 2012; Siegel, 2018). Considering these challenges, the recovery is remarkably quick as the entire U.S. energy sectors pulls together to restore the operational capacity of the electricity grid just 14 to 45 days after the incident. A more comprehensive analysis of the plausibility of such a scenario is provided in Appendix D.

4.2 High-altitude Electromagnetic Pulse

A HEMP could be caused in a variety of ways. For instance, it could be intentionally caused by a small terrorist group that is supported by a nuclear power state such as North Korea (i.e., a malicious risk). It could also intentionally be used by nation states in a conflict situation (i.e., a conflict risk). Finally, it could also be caused unintentionally, for example, in a military accident (i.e., an accident risk).

No matter the cause, researchers have identified increasing vulnerabilities to HEMP and other electromagnetic pulse (EMP) attacks as digital systems and services proliferate in CIs (Savage et al., 2010). As integrated circuits shrink and a wider variety of technologies become reliant on digital microelectronics, the scope of destruction from HEMP attacks has dramatically broadened from a few technologies in the 1980s to the ubiquitous digital systems and services of today. The early time, high frequency component of a (H)EMP known as the E1 pulse component is especially concerning. The minimum electric field of the E1 pulse required to upset the operation of common desktop computers fell by a factor of seven from 1980 to 2001 (Camp & Garbe, 2006). The control center components of SCADA systems are physically similar in design to desktop computers, and are among the most vulnerable technologies to (H)EMP E1 effects. A comprehensive test of several SCADA systems found 100% of the control systems were affected, in many cases observing

[12]However, the recent shutdown of the Colonial Pipeline (a pipeline network which transports a large fraction of the oil on the East Coast of the US) illustrates that a cyberattack could potentially even unintentionally cause CEL (i.e., be an accident risk). In this specific case, low confidence in the cybersecurity measures of the company led to the shutdown of the entire pipeline network after the cybercrime group DarkSide attacked the business network of the company with ransomware (Sanger & Perlroth, 2021).

permanent damage to the system components (Foster et al., 2008). Many SCADA systems communicate using long surface level ethernet cables that can have strong currents induced by an E1 pulse, increasing their vulnerability. In addition, most critical infrastructures, at least in the US, are not shielded against the effects of HEMP with the military branch being the only exception (Caruso, 2015).

For our simulation experiment, we consider a devastating but plausible scenario were an unknown adversarial nation state detonates two HEMP weapons 160 km above the population heavy centers of the West Coast and the East Coast of the U.S. destroying a significant fraction of digital systems such as SCADA systems in the entire country within seconds (Wilson, 2008).[13] As a result of this attack, the electricity grid and multiple other critical infrastructures (e.g., the energy sector at large, information and communications technology sector, etc.) are severely disrupted and only function at about 40% capacity. Due to the large amounts of equipment that need to be replaced on a national scale and potentially extremely long lead times for the replacement of some critical electricity grid components such as high voltage transformers (i.e., several months; U.S. Department of Energy, 2014), we estimate a broad range of recovery times from 60 to 525 days (i.e., 2 month to 1.5 years). A more comprehensive analysis of the plausibility of such a HEMP scenario is provided in Appendix D.

5 Simulation Overview

To investigate the effects of different degrees of digitalization in a society in the aforementioned scenarios, we simulate how the scenarios play out in the context of the U.S. food supply chain. This is a useful choice as the food system is a critical sector of society that has a strong correlation with human well-being. If for whatever reason people start to lose access to food, this directly translates into losses of well-being and, after some time, enhanced probabilities of illness and even death. Thus, loss of access to food can act as reasonable proxy variable to assess the overall severity of a CEL scenario.

Our simulation of the U.S. food supply chain is built on the prior work of Huff et al. (2015). In their work, they have investigated the resilience of the U.S. food system against severe pandemics. For this, they had developed a SD-based simulation focused on modeling the impacts of worker absenteeism on the scale of the entire U.S. on the different stages in the food supply chain, namely, *farms*, *food processing*, *food distribution*, and *food retail*. We built on this work by removing some of the parts focused on modeling the effects of worker absenteeism and replacing them with models of key sectors of society affected by CEL and how they would impact the food supply chain at each stage. The main logic of the resulting model is visualized in Figure 18.[14]

Starting at the top left and moving to the bottom right of Figure 18, the demand flow is moving from the total U.S. population through each stage of the food supply chain down to the farm sector. At each stage the expected demand is stored and used to regulate the production and transportation of food products up to the next stage in the supply chain. Importantly, for each of the major production and transportation steps in the supply chain, the maximum carrying capacity is constrained by the disruption of the key

[13] While we do not consider this, such a scenario could quickly turn into an all-out HEMP exchange with other nuclear power states such as China or Russia (Pry, 2020, 2021) if they are presumed to be responsible for this attack.

[14] Here, we only present the main logic of the model due to the large size and complexity of the complete model. However, the complete model can be retrieved from the Online Appendix and should be reasonably accessible once the main model logic is understood. The model can be run and experimented with using the free VENSIM model reader (https://vensim.com/free-download/).

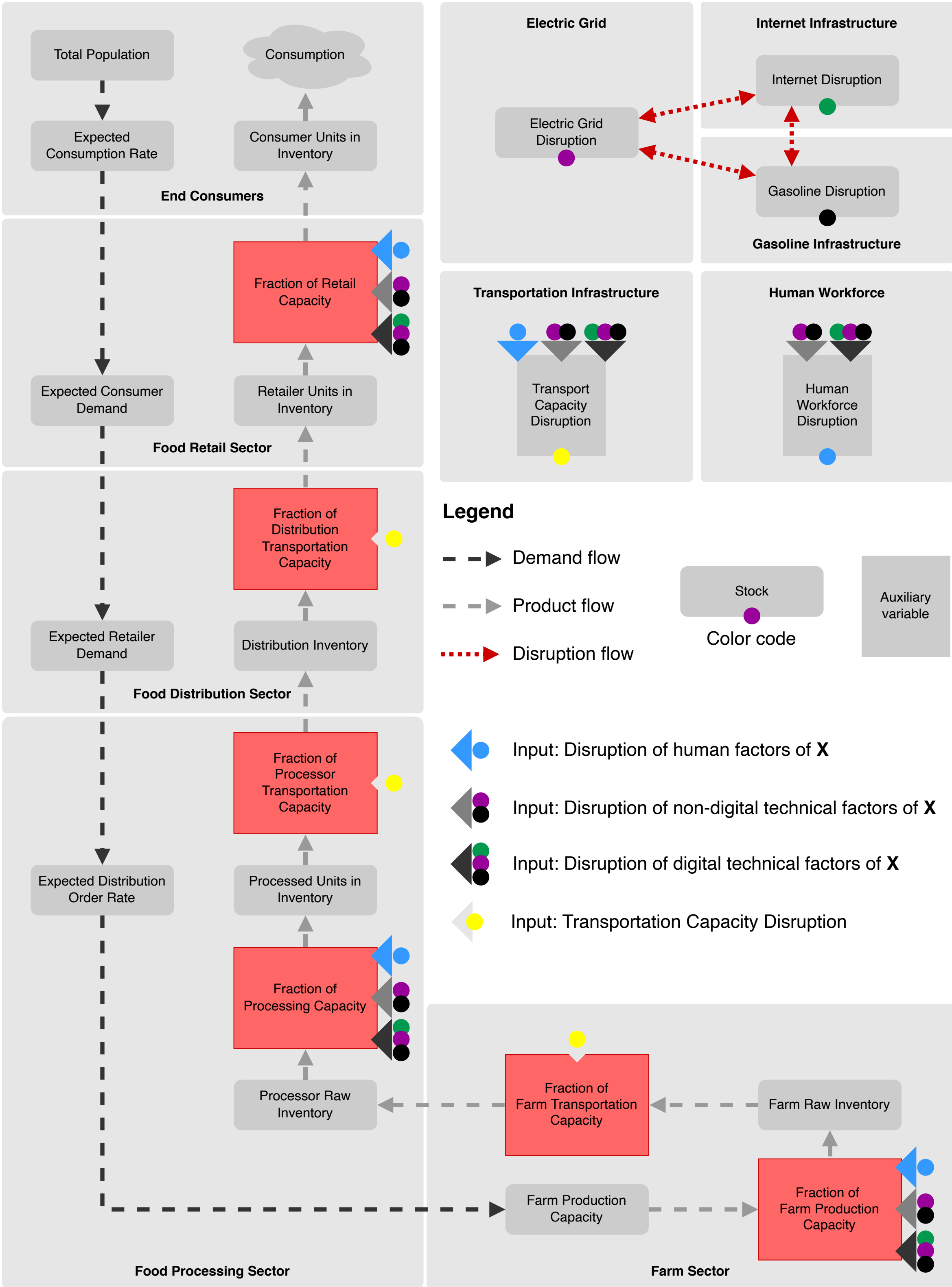

Figure 18: Overview of the Main Logic of the SD-based Simulation Model.

infrastructures the step is dependent on. In particular, we have modeled the constraints in such a way that disruptions to the key infrastructures can interact in a variety of different ways to limit the capacity at each step. Specifically, as part of the modelling, we have taken a socio-technical perspective and distinguished between three broad sources of disruption effects: the *human factors*, the *non-digital technical factors*, and the *digital technical factors* involved in a step. This separation allows us to model the impact of CEL disruptions on a granular level for each step[15] as well as enables us to investigate the impact that different degrees of digitalization have in the case of disruption scenarios.

In terms of the key sources of disruption, we have identified the electricity grid, the Internet, the gasoline infrastructure, the transportation infrastructure and the human workforce as major dependencies of the food system.[16] We explicitly modeled the electricity grid, the Internet, the gasoline infrastructure, and their interdependencies as base infrastructures. The transportation infrastructure and the human workforce are modeled as extending infrastructure that depend on the base infrastructures but do not constrain them in turn.

6 Simulation Experiments

Our goal with this paper is to investigate the effect that different degrees of digitalization in a society would have on the resilience of the food supply chain in the case of CEL scenarios. As a first step toward answering this question, we decided to use our SD-based simulation model to conduct two simulation experiments in two different CEL scenarios. We vary the degree of digitalization in the U.S. food supply chain between one control condition (i.e., current level of digitalization) and one treatment condition (i.e., plausible future higher level of digitalization). For this, we identified 6 parameters in our model that would significantly vary between these two conditions and assigned them plausible value ranges for both conditions as summarized in Table 13.

Table 13: An overview of the model parameters varied between the treatment and control condition of the simulation experiments. See the Online Appendix for a justification of the assigned parameter values.

Location	**Description**	**Control**	**Treatment**
Farm	Overall factor of digital technical factors on total disruption	30% - 50%	50% - 90%
Transport	Overall factor of digital technical factors on total disruption	30% - 70%	70% - 90%
Processing	Overall factor of digital technical factors on total disruption	60% - 80%	80% - 100%
Retail	Overall factor of digital technical factors on total disruption	60% - 80%	80% - 100%

[15]Altogether, we have estimated 56 parameters to model the impact that infrastructure disruptions have on the food supply chain. See the Online Appendix for a detailed documentation.

[16]We acknowledge that the water system is another major dependency. However, due to the strong correlation between the water system and the electricity grid (e.g., Petermann et al., 2011) and in favor of limiting the complexity of the model, we decided against modeling the water system explicitly and instead subsumed its disruption effects within the disruption effects of the electricity grid.

Location	Description	Control	Treatment
Electricity grid	Overall factor of Internet disruption on electricity grid	0% - 20%	20% - 40%
Gasoline infrastructure	Overall factor of Internet disruption on gasoline infrastructure	0% - 20%	20% - 40%

We use the final values of two food system disruption related variables as dependent variables for our experiments: 1) *Peak number of people without access to food,* and 2) *Cumulative number of person days without access to food.* Here, person days without access to food are defined in terms of the difference in the actual rate of consumption and the desired rate of consumption by the population.[17] Both variables together provide a useful summary of how the CEL scenarios and the different degrees of digitalization in societies affect the food supply chain. Whereas the peak number of people without access to food scopes the maximum *intensity* of the disruption, the cumulative number of person days without access to food also reflects the duration of the disruption and, thus, more closely approximates the overall *size* of a disruption.[18]

The two hypotheses to be tested in the simulation experiments are:

Hypothesis 1 *In the case of CEL scenarios, a high degree of digitalization in a society leads to more intense disruptions in the food supply chain.*

Hypothesis 2 *In the case of CEL scenarios, a high degree of digitalization in a society leads to altogether larger disruptions in the food supply chain.*

The testing of these hypotheses was conducted in two simulation experiments that made ease use of two Monte Carlo simulation with 2000 iterations to sample a broad spectrum of the parameter space for the control as well as treatment condition. In particular, all parameter ranges were sampled using a uniform probability distribution reflecting our high uncertainty regarding the real probability distribution of the variables.

6.1 Coordinated Cyberattacks

As we have established in the scenario development section, a coordinated cyberattack on the electricity grid is a plausible and maybe even likely scenario that could lead to CEL. In Table 14, we summarize the scenario-specific parameter values that we estimated for this scenario.

[17] In our calculations of peak numbers of people affected, we naively assume that people either have full or no access to food. While this might not be an entirely realistic assumption, it is reasonable for our simulation experiments as we are not interest in accurate predictions of the number of people affected in CEL but the impact of different degrees of digitalization in societies.

[18] For instance, a high peak number of people without access to food but a comparatively low cumulative number of person days without access to food would indicate an intense but short disruption of the food supply chain, whereas a low peak number of people without access to food but a comparatively high cumulative number of person days without access to food would indicate a less intense but prolonged disruption.

Table 14: An overview of the model parameters varied for the coordinated cyberattacks scenario. See the Online Appendix for a justification of the assigned parameter values.

Location	**Description**	**Value**
Electricity grid	Fraction of primary electricity grid disruption	90%
Electricity grid	Minimum recovery time for disruption	14 – 45 days
Gasoline infrastructure	Fraction of primary gasoline infrastructure disruption	0%
Gasoline infrastructure	Minimum recovery time for disruption	1 – 3 days
Internet	Fraction of primary Internet disruption	0%
Internet	Minimum recovery time for disruption	1 – 2 days

The results of the two Monte Carlo simulation runs for the two key dependent variables in this experiment are visualized as boxplots in Figure 19. Visual inspection of the diagrams suggests, and the statistical analysis of the results confirms, a highly significant positive difference between the control and treatment condition for both dependent variables.[19] In absolute and relative numbers, the difference in outcomes is notable. The mean peak number of people without access to food in the control condition is around 137 million (˜42% of the U.S. population), whereas the mean for the treatment condition is around 183 million (˜55% of the U.S. population; ˜34% increase), and the mean cumulative number of person days without access to food in the control condition is around 3 billion (on average ˜22 days loss of access to food per affected person), whereas the mean for the treatment condition is around 4 billion (˜33% total increase, on average ˜22 days loss of access to food per affected person). Thus, we can accept both hypotheses we investigated for this CEL scenario simulation experiment.

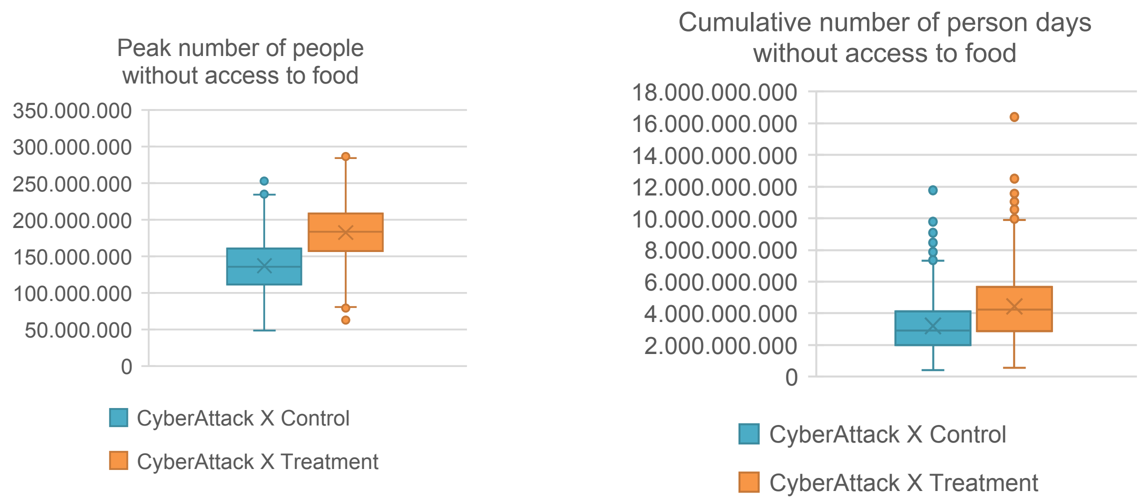

Figure 19: Main Results for the Cyberattack Scenario Simulation Experiment.

[19]Both variables are significantly different between control and treatment condition at 99,9% confidence (Peak number of people without access to food, p-value = 2,2913E-286; Cumulative number of person days without access to food, p-value = 2,09283E-92).

6.2 High-altitude Electromagnetic Pulse

As we have highlighted in the scenario development section, the use of HEMP weapons would likely be devastating for a digital-infused infrastructures and result in very severe CEL. In Table 15, we summarize the scenario-specific parameter values that we estimated for this simulation experiment.

Table 15: An overview of the model parameters varied for the HEMP scenario. See the Online Appendix for a justification of the assigned parameter values.

Location	**Description**	**Value**
Electricity grid	Fraction of primary electricity grid disruption	60%
Electricity grid	Minimum recovery time for disruption	60 – 530 days
Gasoline infrastructure	Fraction of primary gasoline infrastructure disruption	60%
Gasoline infrastructure	Minimum recovery time for disruption	60 – 355 days
Internet	Fraction of primary Internet disruption	60%
Internet	Minimum recovery time for disruption	60 – 355 days

The results of the two Monte Carlo simulation runs for the key dependent variables in this experiment are visualized as boxplots in Figure 20. Again, visual inspection of the diagrams suggests, and the statistical analysis of the results confirms, a highly significant positive difference between the control and treatment condition for both dependent variables.[20] In absolute and relative numbers, the difference in outcomes is again notable. The mean peak number of people without access to food in the control condition is around 172 million (~52% of the U.S. population), whereas the mean for the treatment condition is around 189 million (~57% of the U.S. population; ~10% increase), and the mean cumulative number of person days without access to food in the control condition is around 35 billion (on average ~203 days loss of access to food per affected person), whereas the mean for the treatment condition is around 44 billion (~26% total increase, on average ~233 days loss of access to food per affected person).[21] Thus, we can accept both hypotheses we investigated for this CEL scenario simulation experiment.

7 Discussion

7.1 Uncovering The Risk of Digital Fragility

The simulation experiment results suggest that the ongoing digital transformation of societies may have potentially catastrophic consequences if a CEL scenario would ever come to pass. As demonstrated in our simulation experiments, the digital entanglement and

[20]Both variables are significantly different between control and treatment condition at 99,9% confidence (Peak number of people without access to food, p-value = 2,34197E-27; Cumulative number of person days without access to food, p-value = 1,47921E-21).

[21]It needs to be acknowledged that our model does not consider the possibility of people dying to starvation, which would certainly start to become a possibility and potentially a major driver of disruption dynamics in the case of this scenario given the severity of the simulation results. As such, the numbers should not be taken at face value and only be seen as indicators of relative disruption severity between the treatment and the control condition.

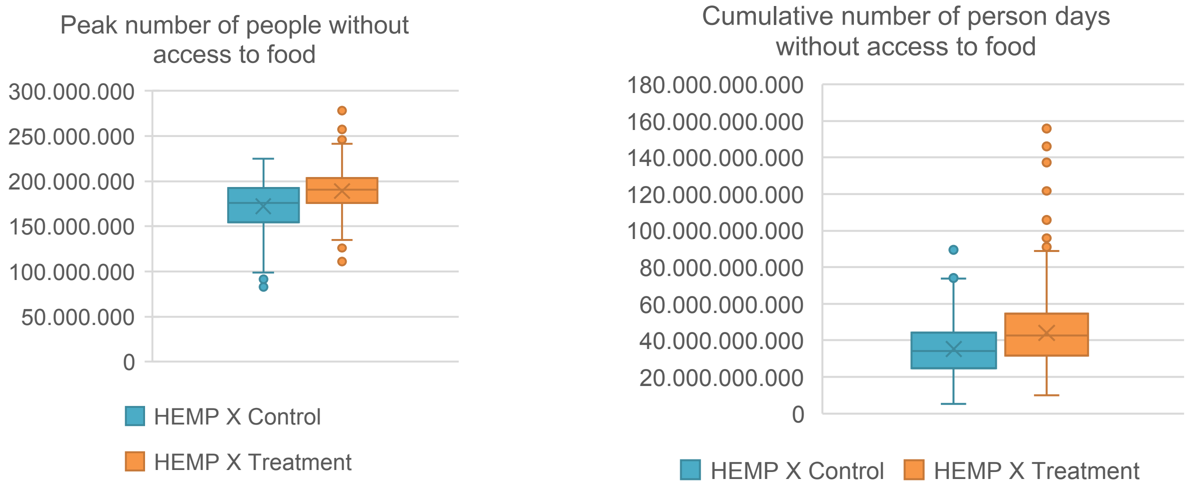

Figure 20: Main Results for the HEMP Scenario Simulation Experiment.

interdependence driven by the digital transformation of societies reinforces and exacerbates catastrophic failure cascades in CEL scenarios, so that ultimately entire countries may be brought down on their knees (Denkenberger et al., 2017; Heino et al., 2019; Petermann et al., 2011; Rinaldi et al., 2001; Stockton & Council, 2016). We suggest to look at such scenarios as being driven by *the risk of digital fragility*: vulnerabilities affecting large fractions of digital systems and services.

As visualized in Figure 21, we find the risk of digital fragility to be brought about by the same features of digital systems and services that also spur the promise of digital agility: standardization and interconnectedness. However, instead of highlighting their potential positive consequences as is the case for the promise of digital agility, the risk of digital fragility highlights their potential negative consequences.

Referring to (1) in Figure 21, the *standardization* of digital systems and services leads to a comparatively homogenous population of technologies, systems and services on the lower levels of the modular-layered architecture of digital technologies (Yoo et al., 2010), which means that some vulnerabilities can affect a large fraction of digital systems. For instance, a bug in a low-level instruction set architecture (ISA) implies that all digital systems using that ISA are vulnerable to that risk (Abu-Ghazaleh et al., 2019). Moreover, as discussed throughout this paper, all digital systems and services are vulnerable to electricity loss. Interestingly, this very risk of exploitable vulnerabilities generally leads to further attempts of standardization around best practices and the most trusted digital systems and services. For instance, there have been multiple attempts to standardize smart grid components to increase the safety of the smart grid (Leszczyna, 2018). However, this reinforcing feedback loop is balanced by the fact that the standardization of digital systems and services also generally enables rapid updating of functions (Yoo et al., 2010). Once discovered, vulnerabilities can often be closed quite quickly. Thus, the risk of digital fragility may also be reduced through the effective and security-minded use of standardization. Nevertheless, this ability to remediate vulnerabilities is not perfect and often limited by situational constraints as illustrated by the slow update cycles associated with a vast majority of Android devices (Mahmoudi & Nadi, 2018).

Regarding (2) in Figure 21, the *interconnectedness* of digital systems and services allows for a rapid propagation of cyber contagions (e.g., a computer virus or worm) or the disruption of only remotely connected digital systems and services, which means that the vulnerabilities which do exist (and are inadvertently triggered or intentionally exploited)

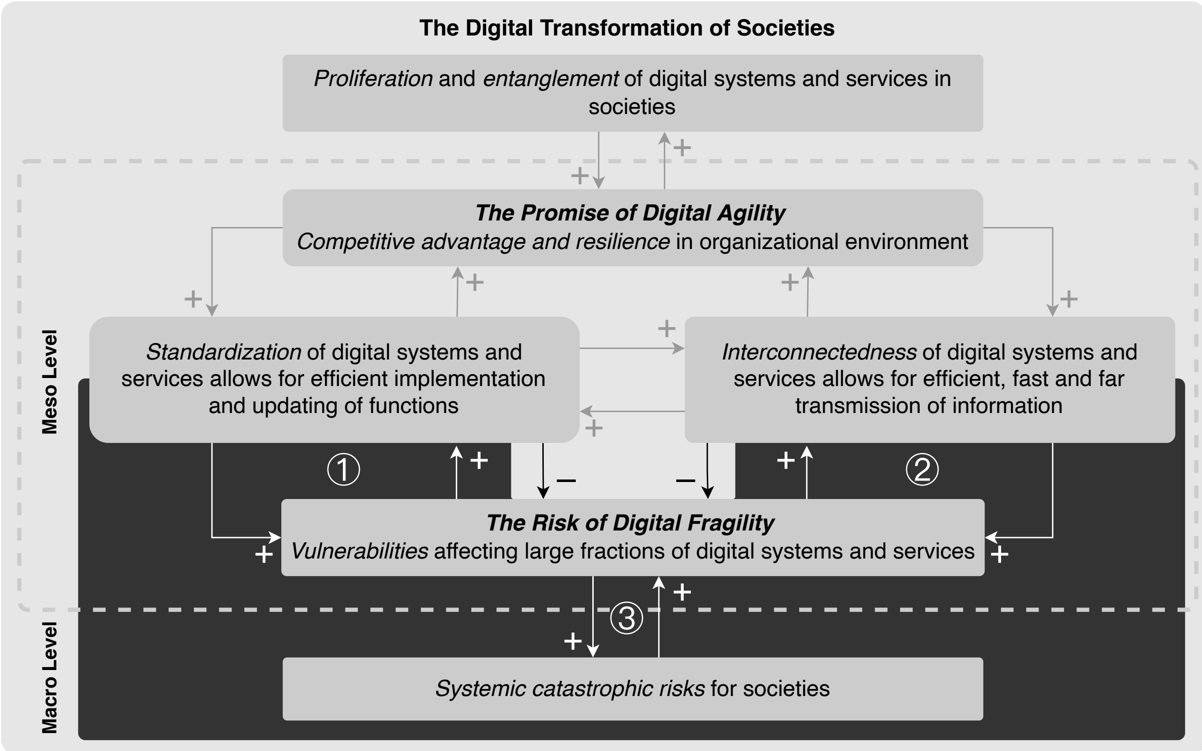

Figure 21: The risk of digital fragility as the driver of dark side of the digital transformation of societies. It describes how competitive dynamics at the center of the digital transformation of societies could cause systemic catastrophic risks. The figure builds on and extends the preceding Figure 16. Numbers label relationships explained in the text.

can lead to massively cascading disruptions. For instance, increasing interconnectedness has so far also tended to contribute to the centralization of information exchange to a few large-scale digital platforms (Evans & Gawer, 2016; Zuboff, 2019). Thus, vulnerabilities in such platforms could have an outsized impact on the rest of the world.[22] In addition, the risks emerging from interconnectedness tend to further strengthen the interconnectedness of digital systems and services as this allows for more adaptive responses to the spreading of disruptions. For instance, the first solution to the increased spreading of computer viruses was not to reduce the interconnectedness of digital systems and services but rather to create antivirus software backed by centralized digital services that further increased the interconnectedness of digital systems and services (Nachenberg, 1997). However, this reinforcing feedback loop is balanced by the fact that interconnectedness of digital systems and services may also allow for the rapid propagation of fixes to vulnerabilities. Extending the antivirus example, if the antivirus system works as advertised the risk of digital fragility may be reduced by the additional connection to the digital services of the antivirus provider. Nevertheless, it must be acknowledged that there are often situational and institutional constraints in place, which limit our collective ability to remove vulnerabilities in an effective and efficient manner (Dunn Cavelty, 2014).

Related to (3) in Figure 21, we argue that these system dynamics contribute to systemic catastrophic risks in the context of the digital transformation of societies because they raise the likelihood for the emergence of disruptive, critical transitions (i.e., catastrophic scenarios) if repeated or large-scale stressors are ever to occur (Scheffer et al., 2012). As network science suggests (Scheffer et al., 2012), highly interconnected networks of

[22] As an example, recently many large websites were taken offline after a vulnerability in the systems of the cloud provider Fastly was accidentally triggered by an unwitting customer (Satariano, 2021).

standardized nodes are able to deal well with small perturbations where local adaptations can distribute the incurred stress in a safe way.[23] However, repeated perturbations or large-scale stressors affecting many nodes can more readily lead to the crossing of a tipping point. As a tipping point is crossed, the capacity for local adaptation is exhausted and catastrophic failure cascades emerge. With the results of our simulation experiments, we have demonstrated this pattern of behavior in CEL scenarios. If the power goes out for an extended period of time, the local adaptive capacities afforded by digital systems and services and the promise of digital agility are overwhelmed by the continuous stress inflicted through electricity loss. A catastrophic failure cascade ensues and sweeps through all interconnected critical sectors of our societies. Importantly, these patterns of behavior are not just limited to CEL scenarios. We suggest that the general dynamics hold irrespective of any specific catastrophic scenario. Loss of electricity is simply a very potent vulnerability that affects all digital systems and services (as well as many non-digital technical systems). As we have discussed in this section, other vulnerabilities that affect large fractions of digital systems and services are poised to emerge over time as the standardization and interconnectedness of digital systems and services continue to increase. As such, the risk of digital fragility may ultimately turn out to be a seriously dangerous Achilles' heel for our societies (Graham, 2011; Manheim, 2020).

7.2 Strategies to Mitigate the Risk of Digital Fragility

In a further step of reflection, we now discuss possible general strategies to mitigate the risk of digital fragility. Our thinking in this section is informed by (a) the available literature on catastrophic risks and emergency preparedness (e.g., Stockton & Council, 2016, 2018), and (b) the general notion of antifragility (Taleb, 2012), which describes generic strategies for overcoming fragility such as creating optionality and redundancy in the system (Derbyshire & Wright, 2014). Altogether, we suggest a tentative set of seven promising strategies to mitigate the risk of digital fragility that provide promising avenues for future research:

- Modularity through protected enclaves;
- Optionality through alternative communication systems and power sources;
- Redundancy through readily available replacement parts;
- Diversity through small-scale experimentation;
- Hormesis through bug bounty programs;
- Better preparedness through financial incentives; and
- Cooperative response through legal obligations.

Modularity Through Protected Enclaves

Increasing the modularity of networks has been shown to reduce the spread of perturbation impact in experimental settings (Gilarranz et al., 2017). Analogously, modularization of CIs through the deliberate construction of protected enclaves has been proposed as an important means of improving our societies resilience to cascading failures (Monken, 2015). For instance, backup facilities should be deliberately located in a way that they have

[23] In the case of the electricity grid, this can be illustrated with the observation that increased interconnectivity between nodes may enable a better balancing of loads and increase resilience in the face of small or random failures in parts of the network (Korkali et al., 2017).

minimal interdependencies with the facilities they ought to replace in case of disaster. One promising avenue for reducing interdependencies in the electricity grid is through loosely coupled microgrids (Hirsch et al., 2018). In the case of cascading failures, microgrids can be operated independently from the main grid, which means that failure cascades can be stopped. Protected enclaves are especially relevant in the case of large-scale disruptions such as CEL scenarios because restoration efforts critically depend on incremental restoration pathways. The most challenging and delicate restoration process is the one which does not have any incremental steps that allow for the bootstrapping of further restoration efforts (c.f. the story of tempus and horus in Simon, 1962). Protected enclaves provide the necessary starting points and foot holes from which to plan and coordinate restoration efforts. However, protected enclaves need to be deliberately designed and developed. IS researchers have demonstrated their ability to develop novel research approaches and solutions that could potentially contribute to the planning, design and development of protected enclaves (e.g., Ketter et al., 2013; Ketter et al., 2016). For instance, the simulation model we developed could be used and extended to further investigate the role and potential positive impact of protected enclaves in the context of the food supply chain.

Optionality Through Alternative Communication Systems and Power Sources

According to Derbyshire and Wright (2014) creating optionality is part of a "truly non-deterministic and non-causal approach to preparation for the future" (p. 220). Having distinct (i.e., largely uncorrelated) options allows to bound potential losses in the face of a broad variety of causes while not limiting potential upsides.

Taking the Internet as an example, it has enabled humanity to coordinate and cooperate on unprecedented scales which has created unprecedented levels of wealth. However, in turn, our systems of organizing have become dependent on the Internet—without it (e.g., in CEL scenarios), we would be crippled (Petermann et al., 2011; Siegel, 2018; Stockton & Council, 2018). Here, the creation of optionality via alternative communication systems that could take over key functions of the Internet but are not tightly linked to the Internet (e.g., they work by different mechanisms) provides a valid strategy to bound the negative effects associated with the loss of the Internet—no matter the cause (Stockton & Council, 2018). Currently, efforts are underway to develop such alternative communication systems based on radio technology and mesh networking (Stockton & Council, 2018). However, given the scale of the task at hand, additional research is likely needed and helpful to create optionality regarding our communication systems (Sakurai et al., 2014). For example, the development of approaches that can integrate existing mobile devices into self-organizing emergency mesh networks could help to create novel and cost effective alternative forms of communication in crisis situations (e.g., Banerjee et al., 2021).

Another key challenge during the onset of catastrophic scenarios such as CEL is maintaining power for lifeline CI, such as drinking water, food, medical services, communication services, and emergency services as well as electrical grid and CI recovery activities (Stockton & Council, 2016). Emergency generators at key facilities provide some availability for lifeline infrastructure; however, the longer a blackout lasts, the higher the likelihood of such generators running out of fuel or failing (Petermann et al., 2011). As such, creating optionality by having additional power sources available in emergency situations seems like a prudent goal. Against this backdrop, hybrid electric vehicles (HEVs) have been identified as useful and cost-effective sources of flexible and mobile power generation, which could help to provide emergency power in crisis situations (Rahimi & Davoudi, 2018; Ustun et al., 2015). IS research could build on such results and investigate novel and cost-effective ways for the organization and use of HEVs in emergency situations. For instance, a smartphone based emergency communication system connected via a mesh network (e.g., Banerjee et al., 2021) could integrate a registry for HEV owners willing to assist during a CEL. Such a system could provide a cost-effective infrastructure for the coordination and deployment of a fleet of portable power generators, potentially greatly improving resilience and recovery efforts in disaster scenarios.

Redundancy Through Readily Available Replacement Parts

Enhancing redundancy and spare capacity is seen as another part of a well-considered strategy to prepare for an inherently uncertain and unpredictable future (Derbyshire & Wright, 2014; Taleb, 2012). While redundancy decreases efficiency, it can be seen as an investment that can provide significant upsides in the case of emergencies (Derbyshire & Wright, 2014). For instance, hospitals and companies with reserves of personal protective equipment or ventilators were well positioned when the COVID-19 pandemic hit, whereas hospitals or companies without reserves quickly found themselves in a precarious situation.

In the context of the CEL scenarios we discussed, redundancy in relation to SCADA systems emerged as potential mitigation strategy. The latest generation of SCADA systems is characterized by their use of commercial off the shelf (COTS) parts, the use of open source software, and a shift in focus from custom hardware to custom software (Good, 2012). While the homogeneity induced by such standardization developments introduces risks, it also opens up a new opportunity for increases in redundancy as defective parts may be more easily replaced with standard computer parts that may have not been damaged as part of a cyberattack or HEMP. Thus, having dedicated stockpiles of standard chips and other computer parts for essential SCADA systems seems like a worthwhile investment in case of emergency. However, how to organize and finance such initiatives is generally an open question as market pressures do not seem to incentivize investments in redundancy and resilience (Little, 2005) as demonstrated by the recent (February 2021) blackouts in Texas, one of the most deregulated electrical grids in the world (Makholm, 2021). Thus, future IS research could look into creative ways to enhance the availability of replacement parts for digital systems in CIs. IS research on the circular economy may provide a fruitful starting point for such endeavors (e.g., Zeiss et al., 2021).

Diversity Through Small-Scale Experimentation

Considering the combination of the modularity, optionality, and redundancy strategies, we arrive at the realization that having a certain level of diversity in a system is desirable as it allows for the evolution of uncorrelated mechanisms that contribute to similar goals and, thus, increase redundancy and resilience.[24] A viable strategy to induce such diversity is the use of small-scale experimentation to establish a diversified portfolio of projects, approaches, and technologies that is conducive to the emergence of redundancy and resilience in a system (Derbyshire & Wright, 2014).

In relation to the risk of digital fragility, small-scale experimentation could take the form of open innovation initiatives (Chesbrough et al., 2006) aimed at developing a variety of projects, approaches, and technologies that could be evaluated in terms of their systemic effects and made use of accordingly. However, open innovation initiatives are still in the early phases of being adopted in the critical sectors of our societies (Greco et al., 2017; Kankanhalli et al., 2017). IS research would seems well positioned to play a leading role in fostering this adoption. For instance, lessons learned from existing open innovation

[24]This thinking is in line with modern portfolio theory (Markowitz, 2010) which argues that investors can reduce their overall risk exposure without reducing returns by investing in a diversified portfolio of assets (i.e., not holding perfectly positively correlated assets).

initiatives (Chesbrough et al., 2006) could be used to help leverage the potential of open innovation for mitigating the risks of digital fragility.

Hormesis Through Bug Bounty Programs

Hormesis generally refers to the biological mechanism by which an organism overcompensates in reaction to the presence of a small dose of toxin and thereby prepares itself for future encounters (Pech & Oakley, 2005). Analogously, in the context of planning for rare and potentially catastrophic events, hormesis stands for the deliberate seeking out of stressors that can help to prepare an organization for worse scenarios (Derbyshire & Wright, 2014; Taleb, 2012).

In relation to digital fragility and CEL scenarios, bug bounty programs (BBPs; Malladi & Subramanian, 2020) seem like a promising means of realizing hormesis in an cost-effective, controllable and safe way. In general, a BBP is understood to be a form of crowdsourcing, where the organizers of the program determine a target system as well as certain rules of engagement and use prizes to incentivize the crowd to find vulnerabilities in the target system (Malladi & Subramanian, 2020). In particular, BBPs can be designed to help identify and fix vulnerabilities in specific products, services or systems and even consider their interdependencies. Moreover, BBPs can be deliberately limited to test environments, where no real harm can occur even if the target system is completely taken over or destroyed. As such, BBPs would lend themselves very well to the identification of specific vulnerabilities in critical sectors of our societies that are introduced by the digital transformation. In particular, BBPs could incentivize an ongoing examination and discovery of unanticipated and potentially catastrophic interdependencies in digital systems and services. IS research would seem to be in a great position to design and steer such efforts, for instance, building on and extending the aforementioned work on open innovation initiatives (Chesbrough et al., 2006).

Better Preparedness Through Financial Incentives

The creation of financial instruments that incentivize disaster preparedness is another potentially effective strategy for mitigating the risks of digital fragility. Here, resilience bonds have been proposed as an instrument to incentivize projects aimed at reducing the negative effects from very severe, relatively rare catastrophes (Vaijhala & Rhodes, 2018). Similar to insurance policies, resilience bonds are paid for in normal times and pay out in the case of disaster. However, in addition to a simple insurance mechanism, probabilistic catastrophe models are leveraged to identify worthwhile risk reduction efforts which may be paid for through discounts to the premium amount. Thus, risk reduction projects could be paid for in a structured and long-term way that is easier to manage and justify for potentially cash-strapped organizations and businesses. Moreover, forward thinking regulation could consider instituting mandatory resilience bonds for CI sectors (similar to banking regulations in the financial sector), which would strongly incentivize more active engagement with cost-effective means of disaster preparedness as a cost cutting measure. Such an incentive structure could potentially overcome the traditional lack of investments into the long-term resilience and sustainability of CIs (Little, 2005). Despite this promising potential, open questions regarding resilience bonds remain, for instance, as of yet the realistic modelling of CI risks and associated risk reduction measures is still viewed as a difficult scientific challenge (Helbing, 2013). As such more work is needed to help establish resilience bonds more broadly in practice. IS research could contribute to such efforts through the design and evaluation of simulation models. For instance, the simulation model developed as part of this paper could be extended and refined to allow for the assessment of potential disaster risk reduction measures in relation to the food supply chain.

Cooperative Response Through Legal Obligations

The final strategy that we consider is to create legally binding agreements for companies that are part of CIs to cooperate more and freely share their resources in the case of extreme disasters. Given the rarity of such events, it would only pose small financial burdens on companies to agree to such legislation but it could prove vital if such events do occur. For example, there already exist agreements that allow cell phones to make emergency calls regardless of subscription status if network coverage is available. Such regulations could be extended to other aspects of CIs. For instance, internet service providers (ISPs) should suspend payment requirements for cyber real estate (i.e., domain names) in countries where a disaster occurs so that information dissemination is not impinged by comparatively trivial technicalities. Another avenue would be to study coopetition strategies during the COVID-19 pandemic (Crick & Crick, 2020) to identify and establish effective mechanisms that support cooperation in disaster scenarios. One challenge that such legal obligations would face is to find effective means of specifying when they become activated as time may be of the essence. For instance, some electric utility providers have contributed to a cascading blackout because they were only willing to act in their narrow self-interest rather than cooperate for the benefit of the stability of the overarching grid (Little, 2005). Finding appropriate governance arrangements and systems that would effectively avoid such situations is an interesting question that future IS research might be able to contribute to. For instance, coordination methods and platforms could be designed and evaluated to guide and inform policy making.

7.3 Implications for Information Systems Research

Given our preceding discussion, what are the main takeaways for IS research? Primarily, we have used two simulation experiments to illustrate the potentially catastrophic effects of the digital transformation of societies in the case of CEL scenarios. The results of our experiments should raise awareness about the risk of digital fragility as an important concern that IS researchers should deliberately address going forward. As stewards of the digital transformation of our societies, we have the moral responsibility to ensure that our societies can continue on (or at least quickly recover) when at some point the lights go out as part of a prolonged blackout. Thus, we believe that IS research should play an important role in the further investigation of the risk of digital fragility as well as in the development of novel solutions that help to mitigate it. We started this endeavor by suggesting seven broad strategies for future work that IS researchers are encouraged to build on.

However, additional steps are necessary. The risk of digital fragility is not a property of an information system that could easily be designed out of the system using a design theory (Gregor & Jones, 2007). Rather it is an ongoing dynamic that needs to be continually managed rather than solved (Ackoff, 1967; Manheim, 2020). This will require systematic research from a variety of fields as well was cooperation and continued well-informed interventions on an unprecedented scale (Helbing, 2013). In many ways IS research seems well positioned to contribute to—or even play a leading role in—such an endeavor, since, as a community, we undoubtedly have a strong history in very relevant socio-technical systems research (Sarker et al., 2019). Thus, we hope that IS research will rise to the challenge and start to consciously address the risk of digital fragility going forward.

7.4 Implications for Information Systems Practice

In terms of implications for IS practice, we mostly want to emphasize the importance of understanding the scope of the potential catastrophic consequences of the risk of digital fragility. It seems fair to say that most IS practitioners today are probably unaware that by promoting the digital transformation of our societies they might be contributing to systemic catastrophic risks that could upend life as we know it. Educating IS practitioners about the risk of digital fragility and ways of mitigating it (e.g., using the seven strategies that we outlined) could become an important concern for IS education. We hope that the framing of the risk of digital fragility as we have presented it in this paper is conducive to this endeavor.

7.5 Implications for Policy Making

Our suggested strategies for mitigating the risk of digital fragility provide important insights for policy makers. Rather than interpreting our warnings as a call to stop investing in the digital transformation of our societies, our strategies suggest that more investments into the deliberate and thoughtful transformation of CIs seem necessary to counterbalance the risk of digital fragility rather than less. A focus on a *differential* digital transformation of societies (i.e., digital transformation directed at advancing resilience and sustainability disproportionately to mere technological capability and efficiency gains) seems important in this regard and could be supported through innovative financial instruments such as resilience bonds (Vaijhala & Rhodes, 2018). Moreover, well thought-out regulation informed by the seven strategies that we have outlined seems like another avenue worthwhile exploring.

8 Limitations and Future Work

Our work is not without limitations. While we believe that the simulation experiments presented in this paper make a rigorous case for the potentially catastrophic consequences of the risk of digital fragility in CEL scenarios, we have to acknowledge that our investigation did not look at the positive outcomes of the promise of digital agility in times without infrastructure disruptions. As such, we cannot make any claims regarding the overall costs and benefits associated with the digital transformation of societies. However, we can highlight that in the case of CEL scenarios we expect a high degree of digitalization to lead to a more intense and overall larger disruption of the food system. This result allows us shine a bright light on a heretofore unrecognized unintended consequence or dark side of the digital transformation of societies. We hope that our work will inspire future work to further investigate the risk of digital fragility and potential mitigation strategies in a variety of scenarios beyond CEL and a diversity of contexts other than the food system. In addition, the SD-based model of the U.S. food supply chain we developed also provides a fruitful foundation for future work. For instance, the model could be used to investigate the impacts of potential mitigation strategies for the risk of digital fragility.

9 Conclusion

In this paper, we have played the devil's advocate to the digital transformation of societies and investigated what would happen when the machines stopped. We have done this by conducting two simulation experiments of extreme but plausible CEL scenarios and how they would affect the U.S. food supply chain. Our results suggest that a high degree of digitalization in a society increases the intensity and size of disruptions in CEL scenarios. We trace this result to the risk of digital fragility, which is induced by the underlying dynamics of the digital transformation of our societies. Informed by extant literature we direct future research to further investigate seven strategies for mitigating the risks of digital fragility. We then summarize key implications for IS research, IS practice, and policy making.

Finally, we want to close this paper with a call to action for the IS research community. As our paper aims to demonstrate, the digital transformation that we are not only researching but often also heavily promoting may have severe unintended consequences we might not be aware of. At the same time, our actions are setting the foundation for future scholars and directing our attention. Let us become aware of this responsibility, broaden our horizons to the possibility of disaster—and then work diligently to prevent it.

> "The price of greatness is responsibility."
>
> — Winston Churchill

Online Appendix

Herwix, A., Tieman, R., Rivers, M., Rosenkranz, C., & Denkenberger, D. (2022). *What Happens When the Machines Stop? Uncovering the Risk of Digital Fragility as the Achilles' Heel of the Digital Transformation of Societies (Online Appendix)*. OSF. https://osf.io/k3c9g/

Appendix A: Literature Review Summary

We have conducted a series of three systematic literature reviews to assess the relevance of prior IS research in relation to the topics of this paper. All literature reviews were carried out on the 25th of May 2022 via the Web of Science database (https://www.webofscience.com) using the eight publications of the Senior Scholar's Basket as a filter. In addition, we used a set of queries to search the topic field (equivalent to searching "title, keyword, and abstract" in other databases). We investigated the results in two steps. First, we read all titles and abstracts to assess potential relevance of the articles. Then we further read into the articles we found broadly relevant to our topic to identify directly related prior work. However, no IS research considering the role of the digital transformation of societies in relation to severe infrastructure disruption scenarios such as blackouts could be found. In particular, we did not find any paper considering the dependency of digital systems and services on electric energy in any detail. See results are summarized in Table 16.

Table 16: Results of the Literature Review

Query	**Total Results**	**Broadly Relevant**	**Directly Related**
catastroph OR disaster* OR crisis OR "extreme event*" OR "black sky" OR blackout* OR "power loss" OR "power outage" OR "electricity loss" OR "product failure*" OR resilience OR fragility*	150	45	0
"dark side" OR "unintended consequence"*	57	2	0
"cyber war" OR "cyber attack" OR "cyber security" OR cybersecurity OR malware*	37	14	0

Appendix B: Simulation Model Validation

Table 17: An overview of the simulation model validation based on Fang et al. (2018).

Test	**What to Test**	**Implementation**
	Direct Structural Tests	
Structure Assessment	Model structure is consistent with relevant descriptive knowledge of the system.	The model is based on the peer-reviewed model of the U.S. food supply chain by Huff et al. (2015). We have also sent our updated version of the model to the original authors for feedback but have not heard back so far.
Parameter Assessment	Parameter values are consistent with and reasonable to descriptive and numerical knowledge of the system.	Judgmental methods were applied by the author team to estimate most of the relevant parameters. Where appropriate, we deferred to the parameter values suggested in Huff et al. (2015). We documented and justified our estimates in the Online Appendix.
Boundary Adequacy	The important concepts for addressing the problem are endogenous to the model.	Key concepts related to the disruption of CIs as well as the degree of digitalization are endogenous in the model.
Dimension Consistency	Each equation is dimensionally consistent without the use of parameters having no real-world meaning.	The model passes the dimension consistency check utility in Vensim PLE+.
	Structure Oriented Behavior Tests	
Extreme Conditions	Model responds plausibly when extreme policies apply.	The model exhibited anticipated behaviors when extreme values were assigned to key constants in the model. A documentation of the behaviors can be found in the Online Appendix.
Sensitivity Test	This is the extent to which numerical values and behaviors change significantly.	The simulation experiment was itself a sensitivity test in the form of two Monte Carlo simulations. The model produced the anticipated behaviors.
Integration Error	Results are not sensitive to the choice of time step or numerical integrating method.	The model was tested with a halving of the time step as well as with, both, the Euler and the Runge-Kutta integration methods.

Test	What to Test	Implementation
	Behavior Tests	
Behavior Reproduction	Model reproduces the behavior of interest in the system.	We used two baseline scenarios of smaller scale disruptions to the electricity grid to assess the appropriateness of the model outputs. We found the outputs to be broadly plausible and in line with expectations. Documentation can be found in the Online Appendix.

Appendix C: Scenario Framework

In this appendix, we present the scenario framework which we used as a backdrop for the development of our scenarios. The framework is adapted from the work of Cotton-Barratt et al. (2020), who conceived of a comprehensive classification of extinction risks for the purpose of defending against them. We found this framework to be the best choice for our purposes because it provided us with an exhaustive overview of general logics by which catastrophic risks could unfold. We have adopted most of the terminology used by Cotton-Barratt et al. (2020) but have made some adjustments to adapt them to the specifics of our use case.[25] In our adapted framework, catastrophic risks are characterized along three dimensions, which may be used to systematically develop plausible CEL scenarios: the *origin*, the *scaling mechanism*, and the *impact mechanism* of the risk.

Origin of a Catastrophic Risk

For a catastrophic risk to become a problem at all, it has to have an *origin* or a beginning. Based on Cotton-Barratt et al.'s (2020) work, Table 18 lists a categorization of catastrophic risk origins in terms of human involvement and intentionality, which together lead to a classification of six risk types for catastrophic shocks caused by humans (i.e., anthropogenic risk) and one type for natural risk:

- *Unseen risks* are those where a few people cause a shock that was unforeseen and unintentional. For example, a critical bug in the Linux kernel disables large parts of the internet.
- *Latent risks* are those where many people together cause a shock that was unforeseen and unintentional. For example, pervasive use of social media creates increasingly resilient filter bubbles which cause a complete breakdown of epistemic security and subsequently social order.
- *Accident risks* are those where a few people cause a shock that was foreseen but unintentional. For example, the unintentional release of a virulent disease causes a pandemic.
- *Commons risks* are those where many people together cause a shock that was foreseen but unintentional. For example, large hurricanes caused by man-made climate change.
- *Malicious risks* are those where a few people cause a shock that was intentional. For example, a coordinated cyberattack by terrorists.
- *Conflict risks* are those where many people together cause a shock that was intentional. For example, multiple HEMP attacks as part of a great power war.
- *Natural risks* are those where a shock is not caused by humans. For example, a 1 in 10,000 year solar storm disables a significant fraction of extra high voltage (EHV) transformers globally.

[25] In particular, we reframe extinction risk in terms of catastrophic risks because we suggest that the classification of Cotton-Barratt et al. (2020) can be usefully applied to smaller scales than human extinction if a catastrophic risk is interpreted not in an absolute sense but relative to a focal system under investigation. For instance, on the one hand, given a company as the focal system a catastrophic risk might be seen as anything that causes the breakdown of the company. On the other hand, given the global ecosystem of CIs as the focal system a catastrophic risk is more similar to what Cotton-Barratt et al. (2020) called extinction risk.

Table 18: A categorization of risk types according to their origin based on Cotton-Barratt et al. (2020).

		Anthropogenic Risk		Natural Risk
		Few People Cause Shock	**Many People Cause Shock**	**No People Cause Shock**
Unintentional Harm	**Unforeseen Harm**	*Unseen Risk*	*Latent Risk*	*Natural Risk*
	Foreseen Harm	*Accident Risk*	*Commons Risk*	
Intentional Harm		*Malicious Risk*	*Conflict Risk*	

Scaling Mechanism of a Catastrophic Risk

Once a catastrophic risk has begun to unfold, it has to reach a certain scale to substantially start disrupting and breaking down the functioning of the focal system. According to Cotton-Barratt et al. (2020), it is useful to characterize the process of disruption along two dimensions: (a) the amount of disruption done before a response is possible, and (b) the largest one-step increase in disruption. Based on those dimensions they suggest a classification of three different risk types (see Table 19):

- *Leverage risks* are those where only a low amount of disruption is caused before a response is possible but a large one-step increase in disruption can be observed. For example, a nuclear attack with intercontinental missiles could be detected early without much disruption caused but disruption would rapidly escalate once the missile reached its target.
- *Cascading risks* are those where only a low amount of disruption is caused before a response is possible and the one-step increases in disruption remain small. For example, a virus which only ever causes small amounts of disruptions on the micro-level but could quickly cascade to large-scale disruptions due to self-propagating and exponential growth.
- *Large risks* are those where already a large amount of disruption is caused before a response is possible. For example, a gamma ray burst from a source close to earth could cause irreparable disruption to CIs on a global scale within seconds.

Table 19: A categorization of risk types according to their scaling mechanism based on Cotton-Barratt et al. (2020).

		Amount of disruption before we can respond	
		Low	**High**
Largest one-step increase in disruption	**High**	Leverage Risk	Large Risk
	Low	Cascading Risk	

Impact Mechanism of a Catastrophic Risk

Finally, it is possible to classify catastrophic risks by the mechanism with which they impact the functioning of the focal system. Inspired by Cotton-Barratt et al. (2020), we suggest three risk types based on whether they directly or indirectly affect the focal system (see Table 20):

- *Functioning risks* are those where the focal system's ability to carry out behaviors that fulfill desired functions is directly impacted. For example, a hurricane physically impacts the electricity grid.
- *Infrastructure risks* are those where the focal system is indirectly impacted by failures in the supporting infrastructure. For example, a disruption of the internet impacts the proper functioning of the energy markets.
- *Environment risks* are those where the focal system is indirectly impacted by shocks or changes in the environment. For example, after the nuclear accident in Fukushima the political climate around nuclear power changed dramatically, which led to the early retirement of otherwise well working nuclear power plants.

Table 20: A categorization of risk types according to their impact mechanism.

	Impact		
	Direct	**Indirect**	
Focal System	*Functioning Risk*	*Infrastructure Risk*	*Environment Risk*

Appendix D: Plausibility of the Scenarios

Plausibility of the Coordinated Cyberattack Scenario

The potential for cyber threats to impact electricity infrastructure was first uncovered during the 'Aurora experiment' during which a replica power plant's control systems were hacked causing it to self-destruct (Mackinnon et al., 2013). The ability for cyber risks to cause physical damage to CIs was later realized when the Stuxnet computer worm damaged a uranium enrichment facility in Iran in 2009 – 2010 (Nicolas et al., 2011). Stuxnet was likely able to infiltrate air gapped networks by infecting USB devices that crossed the air gap. It then exploited vulnerabilities in SCADA systems, infecting programmable logic controllers (PLCs) involved in controlling gas centrifuge speed. By covertly fluctuating rotation speeds, Stuxnet was able to destroy hundreds of gas centrifuges. The Stuxnet worm was so infectious that it has been found in control systems of various CI networks outside of Iran, including Chevron's network in 2010 (R. King, 2012; Nicolas et al., 2011). The targeting of SCADA system components by Stuxnet demonstrates the vulnerability of CI processes to cyberattacks.

The increasing connectivity of SCADA systems and utilization of distributed architectures increases the risk of cyberattacks. Wireless SCADA systems using the Internet are increasing in popularity (Pliatsios et al., 2020). As opposed to having to physically interact with hardwired SCADA systems, the increasingly commonplace connection of SCADA systems to corporate networks and even the Internet means that remote attacks should be expected (Kang et al., 2014; Nicholson et al., 2012). Coupled with the utilization of standard protocols to streamline control processes the likelihood of process-aware attacks[26] is increasing (Khorrami et al., 2016). A similar method as the Stuxnet worm may be used to infiltrate SCADA systems, hack into gateways or edge devices, and take control of entire CI nodes. Reprogramming the SCADA system would allow maloperation or self-destruction (Stockton & Council, 2016).

Thus, it is entirely plausible and prudent to be concerned about a coordinated cyberattack that can take down the electric grid by attacking SCADA systems (Onyeji et al., 2014). For instance, there already exist multiple instances of the hacking of SCADA systems in electrical infrastructure (Byres et al., 2007; Kuvshinkova, 2003). Such an attack is of special concern because it could be widespread and have long-term destructive consequences.

Outside of SCADA system vulnerabilities, there exist other ways the electrical grid could be attacked. A recent example is the shutdown of a large pipeline network on the East Coast of the US due to a ransomware attack by the cybercrime group DarkSide (Sanger & Perlroth, 2021). In this cyberattack only the business network of the pipeline company was attacked but due to low confidence in the cybersecurity measures of the company, even operational systems had to be shut. Thus, even relatively simple cyberattacks without the intent to cause catastrophic harm can cascade to become significant events with potentially devastating consequences. If the shutdown had persisted for only a few days longer, it is very likely that even more severe cascade effects across several CIs would have started to appear (Sanger & Perlroth, 2021).

Disconcertingly, the risk of such events may further increase due to the accelerated digital transformation of business practices in response to the COVID-19 pandemic. In particular, an increase in remote work facilitated by greater utilization of digital technologies can be observed (Dwivedi et al., 2020; Papadopoulos et al., 2020), which presents new

[26] Attacks which change run-time parameters or control logic in computational nodes across multiple similar processes.

vectors that malicious actors could potentially exploit, in order to gain information, or interfere with processes. Already a strong increase in damages caused by cybercrime can be observed (IC3, 2021), even with deaths being attributed to ransomware attacks on hospitals (Wolff, 2020).

Given this climate of increasingly prevalent cybercrime, a troubling development is that emerging smart grids appear to be particularly vulnerable to cyberattacks due to increased reliance on communication networks to provide enhanced efficiency and reliability (Wang & Lu, 2013). Moreover, even non-smart electrical grids are vulnerable to cyberattacks. For instance, security researchers have demonstrated how botnets[27] could push non-smart electricity grids into an unstable state by modulating the power consumption of hacked devices in a coordinated way (Dabrowski et al., 2017; Soltan & Mittal, 2018).

Altogether, we conclude that current trends and developments make a CEL scenario due to a coordinated cyberattack at least plausible or even likely in the future.

Plausibility of the High-altitude Electromagnetic Pulse Scenario

On July 9th, 1962 a 1.4 Megaton of TNT equivalent test nuclear warhead known as "Starfish Prime" was detonated 400 km above a remote region in the Pacific, causing a HEMP. 750 miles away and seconds after the blast, telephone communications failed between Kauai, Hawaii and the rest of Hawaii. 900 miles away in Oahu, Hawaii, hundreds of street lights failed, car ignition systems were fused, and high frequency radio equipment was damaged. Based on publicly available information, between a few days and 6 months after the attack most satellites failed due to the explosion (National Coordinating Center for Communications, 2019).

While no HEMPs have been tested since the nuclear test ban treaty of 1963, HEMPs may still be used today by a nuclear enabled nation-state to disable the industry of one or many adversary nation-states. For instance, there have been indications that at least one country (Iran) has been practicing for a HEMP attack (Wilson, 2008) and China as well as Russia consider HEMP attacks as part of the arsenal of modern information warfare with lower thresholds for use than traditional nuclear attacks (Pry, 2020, 2021). Nuclear proliferation, the increased reliance on sensitive digital components for adversary nations' CIs, and the possibility that a HEMP might not induce the deterrence of mutually assured destruction (MAD) increase the plausibility of near-term HEMP attacks (Wilson, 2008). While research estimating the likelihood of a HEMP attack is lacking, previous work estimates a probability of 0.3% per year for full-scale nuclear war (Denkenberger & Pearce, 2018). We suspect the probability of a HEMP attack is likely as high or higher than full-scale nuclear war, due to game theoretic considerations concerning whether HEMP would elicit a full nuclear response. Thus, we conclude that current trends and developments make a CEL scenario due to a HEMP a hopefully unlikely but still plausible scenario in the future.

[27] A botnet is a network of a large number of hacked devices which is under the control of an adversary.

References

Abdelkhaliq, M., Denkenberger, D., Cole, D., Griswold, M., Pearce, J., & Taylor, A. R. (2016). Non Food Needs if Industry is Disabled. *Proceedings of the 6th International Disaster and Risk Conference*.

Abu-Ghazaleh, N., Ponomarev, D., & Evtyushkin, D. (2019). How the spectre and meltdown hacks really worked. *IEEE Spectrum*, *56*(3), 42–49. https://doi.org/10.1109/MSPEC.2019.8651934

Ackoff, R. L. (1967). Management misinformation systems. *Management science*, *14*(4), B147–B156.

Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures*, *46*, 23–40. https://doi.org/10.1016/j.futures.2012.10.003

Banerjee, I., Warnier, M., Brazier, F. M. T., & Helbing, D. (2021). Introducing participatory fairness in emergency communication can support self-organization for survival. *Scientific Reports*, *11*(1), 7209. https://doi.org/10.1038/s41598-021-86635-y

Barley, W. C., Leonardi, P. M., & Bailey, D. E. (2012). Engineering Objects for Collaboration: Strategies of Ambiguity and Clarity at Knowledge Boundaries. *Human Communication Research*, *38*(3), 280–308. https://doi.org/10.1111/j.1468-2958.2012.01430.x

Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., & Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*, *464*(7291), 1025–1028. https://doi.org/10.1038/nature08932

Byres, E., Leversage, D., & Kube, N. (2007). Security incident and trends in SCADA and process industries: A statistical review of the Industrial Security Incident Database (ISID). *White Paper, Symantec Corporation, Cupertino, California*.

Camp, M., & Garbe, H. (2006). Susceptibility of Personal Computer Systems to Fast Transient Electromagnetic Pulses. *IEEE Journals & Magazine*. Retrieved November 27, 2020, from https://ieeexplore.ieee.org/abstract/document/4014644

Caruso, M. (2015). The Threat: The State of Preparedness Against the Threat of an Electromagnetic Pulse (EMP) Event. Retrieved May 28, 2021, from https://oversight.house.gov/wp-content/uploads/2015/05/Caruso-Statement-5-13-EMP.pdf

Centeno, M. A., Nag, M., Patterson, T. S., Shaver, A., & Windawi, A. J. (2015). The Emergence of Global Systemic Risk. *Annual Review of Sociology*, *41*(1), 65–85. https://doi.org/10.1146/annurev-soc-073014-112317

Chang, S. E., McDaniels, T. L., Mikawoz, J., & Peterson, K. (2007). Infrastructure failure interdependencies in extreme events: Power outage consequences in the 1998 Ice Storm. *Natural Hazards*, *41*(2), 337–358. https://doi.org/10.1007/s11069-006-9039-4

Chesbrough, H., Vanhaverbeke, W., & West, J. (2006, August 10). *Open Innovation: Researching a New Paradigm*. OUP Oxford.

Cole, D. D., Denkenberger, D., Griswold, M., Abdelkhaliq, M., & Pearce, J. (2016). Feeding Everyone if Industry is Disabled. *IDRC DAVOS 2016 Integrative Risk Management*

- *Towards Resilient Cities*. Retrieved August 15, 2019, from https://hal.archives-ouvertes.fr/hal-02113486

Cotton-Barratt, O., Daniel, M., & Sandberg, A. (2020). Defence in Depth Against Human Extinction: Prevention, Response, Resilience, and Why They All Matter. *Global Policy*, *11*(3), 271–282. https://doi.org/10.1111/1758-5899.12786

Crick, J. M., & Crick, D. (2020). Coopetition and COVID-19: Collaborative business-to-business marketing strategies in a pandemic crisis. *Industrial Marketing Management*, *88*, 206–213. https://doi.org/10.1016/j.indmarman.2020.05.016

Dabrowski, A., Ullrich, J., & Weippl, E. R. (2017). Grid Shock: Coordinated Load-Changing Attacks on Power Grids: The Non-Smart Power Grid is Vulnerable to Cyber Attacks as Well. *Proceedings of the 33rd Annual Computer Security Applications Conference*, 303–314. https://doi.org/10.1145/3134600.3134639

D'Arcy, J., Gupta, A., Tarafdar, M., & Turel, O. (2012). Reflecting on the "Dark Side" of Information Technology Use. *Communications of the Association for Information Systems*, *35*, 5. https://doi.org/10.17705/1cais.03505

Denkenberger, D. C., & Pearce, J. M. (2018). Cost-effectiveness of interventions for alternate food in the United States to address agricultural catastrophes. *International journal of disaster risk reduction*, *27*, 278–289. https://doi.org/10.1016/j.ijdrr.2017.10.014

Denkenberger, D. C., Cole, D. D., Abdelkhaliq, M., Griswold, M., Hundley, A. B., & Pearce, J. M. (2017). Feeding everyone if the sun is obscured and industry is disabled. *International Journal of Disaster Risk Reduction*, *21*, 284–290. https://doi.org/10.1016/j.ijdrr.2016.12.018

Derbyshire, J., & Wright, G. (2014). Preparing for the future: Development of an 'antifragile' methodology that complements scenario planning by omitting causation. *Technological Forecasting and Social Change*, *82*, 215–225. https://doi.org/10.1016/j.techfore.2013.07.001

Dobson, I., Carreras, B. A., Lynch, V. E., & Newman, D. E. (2007). Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *17*(2), 026103. https://doi.org/10.1063/1.2737822

Dolan, T. E. (2018, February 26). *The London Black Sky Seminar 2018: Infrastructure and Societal Resilience to Black Sky Hazards*. Washington DC, USA, EIS (Electric Infrastructure Security) Council. Retrieved September 25, 2020, from https://www.eiscouncil.org/App_Data/Upload/3f6282a1-4b27-498c-b4c6-c67219d75a30.pdf

Dong, J. Q. (2022). Using Simulation in Information Systems Research. *Journal of the Association for Information Systems*, *23*(2), 408–417. https://doi.org/10.17705/1jais.00743

Dopfer, K., Foster, J., & Potts, J. (2004). Micro-meso-macro. *Journal of Evolutionary Economics*, *14*(3), 263–279. https://doi.org/10.1007/s00191-004-0193-0

Dunn Cavelty, M. (2014). Breaking the Cyber-Security Dilemma: Aligning Security Needs and Removing Vulnerabilities. *Science and Engineering Ethics*, *20*(3), 701–715. https://doi.org/10.1007/s11948-014-9551-y

Dwivedi, Y. K., Hughes, D. L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J. S., Gupta, B., Lal, B., Misra, S., Prashant, P., Raman, R., Rana, N. P., Sharma, S. K., & Upadhyay, N. (2020). Impact of COVID-19 pandemic on information management research and practice: Transforming education, work and life. *International Journal of Information Management*, *55*, 102211. https://doi.org/10.1016/j.ijinfomgt.2020.102211

Evans, P. C., & Gawer, A. (2016). *The Rise of the Platform Enterprise: A Global Survey*. The Center for Global Enterprise.

Fang, Y., Lim, K., Qian, Y., & Feng, B. (2018). System Dynamics Modeling for Information Systems Research: Theory of Development and Practical Application. *Management Information Systems Quarterly*, *42*(4), 1303–1329.

Foster, J., Gjelde, E., Graham, W. R., Hermann, R. J., Kluepfel, H. M., Lawson, R. L., Soper, G. K., Wood, L. L., & Woodard, J. B. (2008, April 1). *Report of the Commission to Assess the Threat to the United States from Electromagnetic Pulse (EMP) Attack: Critical National Infrastructures*. Retrieved November 27, 2020, from https://apps.dtic.mil/sti/citations/ADA484672

Gao, J., Buldyrev, S. V., Stanley, H. E., & Havlin, S. (2012). Networks formed from interdependent networks. *Nature Physics*, *8*(1), 40–48. https://doi.org/10.1038/nphys2180

Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2022). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, *31*(3), 410–435. https://doi.org/10.1080/0960085X.2021.1927213

Gilarranz, L. J., Rayfield, B., Liñán-Cembrano, G., Bascompte, J., & Gonzalez, A. (2017). Effects of network modularity on the spread of perturbation impact in experimental metapopulations. *Science*, *357*(6347), 199–201. https://doi.org/10.1126/science.aal4122

Gladwell, M. (2000). *The tipping point: How little things can make a big difference* (1st ed). Little, Brown.

Good, J. (2012). *Blackstarting the North American power grid after a nuclear electromagnetic pulse (EMP) event or major solar storm* (Doctoral dissertation). James Madison University. Retrieved November 27, 2020, from https://www.semanticscholar.org/paper/Blackstarting-the-North-American-power-grid-after-a-Good/55c6a413e94ccb0523ed17e9002d10312c590306

Goodwin, P., & Wright, G. (2010). The limits of forecasting methods in anticipating rare events. *Technological Forecasting and Social Change*, *77*(3), 355–368. https://doi.org/10.1016/j.techfore.2009.10.008

Graham, S. (2011). Disrupted Cities: Infrastructure Disruptions as the Achilles Heel of Urbanized Societies, 15.

Greco, M., Locatelli, G., & Lisi, S. (2017). Open innovation in the power & energy sector: Bringing together government policies, companies' interests, and academic essence. *Energy Policy*, *104*, 316–324. https://doi.org/10.1016/j.enpol.2017.01.049

Gregor, S., & Jones, D. (2007). The Anatomy of a Design Theory. *Journal of AIS*, *8*(5), 312–335.

Haes Alhelou, H., Hamedani-Golshan, M., Njenda, T., & Siano, P. (2019). A Survey on Power System Blackout and Cascading Events: Research Motivations and Challenges. *Energies*, *12*(4), 682. https://doi.org/10.3390/en12040682

Haigh, T. (2022). Becoming universal. *Communications of the ACM*, *65*(2), 25–30. https://doi.org/10.1145/3506578

Hanelt, A., Bohnsack, R., Marz, D., & Antunes, C. (2021). A systematic review of the literature on digital transformation: Insights and implications for strategy and organizational change. *Journal of Management Studies*, *58*(5), 1159–1197. https://doi.org/10.1111/joms.12639

Heino, O., Takala, A., Jukarainen, P., Kalalahti, J., Kekki, T., & Verho, P. (2019). Critical Infrastructures: The Operational Environment in Cases of Severe Disruption. *Sustainability*, *11*(3), 838. https://doi.org/10.3390/su11030838

Helbing, D. (2013). Globally networked risks and how to respond. *Nature*, *497*(7447), 51–59. https://doi.org/10.1038/nature12047

Herwix, A., & Haj-Bolouri, A. (2021). Revisiting the Problem of the Problem – An Ontology and Framework for Problem Assessment in IS Research. *Proceedings of the Twenty-Nineth European Conference on Information Systems (ECIS2021)*.
Hillis, D. (2010). The Age of Digital Entanglement. *Scientific American*, *303*(3), 93–93.
Hirsch, A., Parag, Y., & Guerrero, J. (2018). Microgrids: A review of technologies, key drivers, and outstanding issues. *Renewable and Sustainable Energy Reviews*, *90*, 402–411. https://doi.org/10.1016/j.rser.2018.03.040
Hollick, M., & Katzenbeisser, S. (2019). Resilient Critical Infrastructures. In C. Reuter (Ed.), *Information technology for peace and security: It applications and infrastructures in conflicts, crises, war, and peace* (pp. 305–318). Springer Fachmedien. https://doi.org/10.1007/978-3-658-25652-4_14
Huff, A. G., Beyeler, W. E., Kelley, N. S., & McNitt, J. A. (2015). How resilient is the United States' food system to pandemics? *Journal of Environmental Studies and Sciences*, *5*(3), 337–347. https://doi.org/10.1007/s13412-015-0275-3
IC3. (2021). *Internet Crime Report 2020*. FBI. Retrieved May 28, 2021, from https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf
Kahn, H., & Wiener, A. J. (1967). *The year 2000; a framework for speculation on the next thirty-three years*. New York, Macmillan. Retrieved May 17, 2021, from http://archive.org/details/year2000framewor00kahn
Kang, D., Kim, B., & Na, J. (2014). Cyber threats and defence approaches in SCADA systems. *16th International Conference on Advanced Communication Technology*, 324–327. https://doi.org/10.1109/ICACT.2014.6778974
Kankanhalli, A., Zuiderwijk, A., & Tayi, G. K. (2017). Open innovation in the public sector: A research agenda. *Government Information Quarterly*, *34*(1), 84–89. https://doi.org/10.1016/j.giq.2016.12.002
Katina, P. F., & Keating, C. B. (2015). Critical infrastructures: A perspective from systems of systems. *International Journal of Critical Infrastructures*, *11*(4), 316–344. https://doi.org/10.1504/IJCIS.2015.073840
Kenett, D. Y., Perc, M., & Boccaletti, S. (2015). Networks of networks – An introduction. *Chaos, Solitons & Fractals*, *80*, 1–6. https://doi.org/10.1016/j.chaos.2015.03.016
Ketter, W., Collins, J., & Reddy, P. (2013). Power TAC: A competitive economic simulation of the smart grid. *Energy Economics*, *39*, 262–270. https://doi.org/10.1016/j.eneco.2013.04.015
Ketter, W., Peters, M., Collins, J., & Gupta, A. (2016). Competitive BENCHMARKING: AN IS RESEARCH APPROACH TO ADDRESS WICKED PROBLEMS WITH BIG DATA AND ANALYTICS. *MIS Quarterly*, *40*(4), 34.
Khorrami, F., Krishnamurthy, P., & Karri, R. (2016). Cybersecurity for Control Systems: A Process-Aware Perspective. *IEEE Design Test*, *33*(5), 75–83. https://doi.org/10.1109/MDAT.2016.2594178
King, A., & Gallagher, M. (2020). *Cyberspace Solarium Commission*. Retrieved May 28, 2021, from https://www.solarium.gov/
King, R. (2012). Virus Aimed at Iran Infected Chevron Network [newspaper]. *Wall Street Journal*. Retrieved November 11, 2020, from https://online.wsj.com/article/SB10001424127887324894104578107223667421796.html
Korkali, M., Veneman, J. G., Tivnan, B. F., Bagrow, J. P., & Hines, P. D. H. (2017). Reducing Cascading Failure Risk by Increasing Infrastructure Network Interdependence. *Scientific Reports*, *7*(1), 44499. https://doi.org/10.1038/srep44499
Kuvshinkova, S. (2003). Sql Slammer worm lessons learned for consideration by the electricity sector. *North American Electric Reliability Council*, *1*(2), 5.

Leszczyna, R. (2018). A review of standards with cybersecurity requirements for smart grid. *Computers & Security*, *77*, 262–276. https://doi.org/10.1016/j.cose.2018.03.011

Little, R. G. (2005). Tending the infrastructure commons: Ensuring the sustainability of our vital public systems. *Structure and Infrastructure Engineering*, *1*(4), 263–270. https://doi.org/10.1080/15732470500103708

Mackinnon, L., Bacon, L., Gan, D., Loukas, G., Chadwick, D., & Frangiskatos, D. (2013, March 1). Cyber Security Countermeasures to Combat Cyber Terrorism. https://doi.org/10.1016/B978-0-12-407191-9.00020-X

Maher, T., & Baum, S. (2013). Adaptation to and Recovery from Global Catastrophe. *Sustainability*, *5*(4), 1461–1479. https://doi.org/10.3390/su5041461

Mahmoudi, M., & Nadi, S. (2018). The Android update problem: An empirical study. *Proceedings of the 15th International Conference on Mining Software Repositories*, 220–230. https://doi.org/10.1145/3196398.3196434

Makholm, J. D. (2021). The Texas Energy Debacle and the Economists. *Climate and Energy*, *37*(10), 19–25. https://doi.org/10.1002/gas.22229

Malladi, S. S., & Subramanian, H. C. (2020). Bug Bounty Programs for Cybersecurity: Practices, Issues, and Recommendations. *IEEE Software*, *37*(1), 31–39. https://doi.org/10.1109/MS.2018.2880508

Manheim, D. (2020). The Fragile World Hypothesis: Complexity, Fragility, and Systemic Existential Risk. *Futures*, *122*, 102570. https://doi.org/10.1016/j.futures.2020.102570

Markowitz, H. M. (2010). Portfolio Theory: As I Still See It. *Annual Review of Financial Economics*, *2*(1), 1–23. https://doi.org/10.1146/annurev-financial-011110-134602

Mikalef, P., Conboy, K., Lundstrom, J. E., & Popovic, A. (2022). Thinking responsibly about responsible ai and 'the dark side' of ai. *European Journal of Information Systems*, *31*(3), 257–268. https://doi.org/10.1080/0960085X.2022.2026621

Monken, J. (2015). Black sky: Exposing electricity as the Achilles' heel of resilience. *Journal of Business Continuity & Emergency Planning*, *9*(1), 25–30.

Nachenberg, C. (1997). Computer virus-antivirus coevolution. *Communications of the ACM*, *40*(1), 46–51. https://doi.org/10.1145/242857.242869

National Coordinating Center for Communications. (2019). Electromagnetic Pulse (EMP) Protection and Resilience Guidelines for Critical Infrastructure and Equipment. (2.2). https://www.cisa.gov/sites/default/files/publications/19_0307_CISA_EMP-Protection-Resilience-Guidelines.pdf

NERC. (2012, May 9). *Severe Impact Resilience: Considerations and Recommendations*. North American Electric Reliability Corporation. Retrieved October 24, 2020, from https://www.ourenergypolicy.org/wp-content/uploads/2012/05/SIRTF_Final_May_9_2012-Board_Accepted.pdf

Nicholson, A., Webber, S., Dyer, S., Patel, T., & Janicke, H. (2012). Scada security in the light of Cyber-Warfare. *Computers & Security*, *31*(4), 418–436. https://doi.org/10.1016/j.cose.2012.02.009

Nicolas, F., Liam, O. M., & Eric, C. (2011, February). *W32. Stuxnet Dossier* (Version 1.4). Retrieved November 11, 2020, from https://docs.broadcom.com/doc/security-response-w32-stuxnet-dossier-11-en

Onyeji, I., Bazilian, M., & Bronk, C. (2014). Cyber Security and Critical Energy Infrastructure. *The Electricity Journal*, *27*(2), 52–60. https://doi.org/10.1016/j.tej.2014.01.011

Papadopoulos, T., Baltas, K. N., & Balta, M. E. (2020). The use of digital technologies by small and medium enterprises during COVID-19: Implications for theory and practice. *International Journal of Information Management*, *55*, 102192. https://doi.org/10.1016/j.ijinfomgt.2020.102192

Parandehgheibi, M., & Modiano, E. (2013). Robustness of interdependent networks: The case of communication networks and the power grid. *2013 IEEE Global Communications Conference (GLOBECOM)*, 2164–2169. https://doi.org/10.1109/GLOCOM.2013.6831395

Pech, R. J., & Oakley, K. E. (2005). Hormesis: An evolutionary "predict and prepare" survival mechanism. *Leadership & Organization Development Journal*, *26*(8), 673–687. https://doi.org/10.1108/01437730510633737

Petermann, T., Bradke, H., Lüllmann, A., Poetzsch, M., & Riehm, U. (2011). *What happens during a blackout? Consequences of a prolonged and wide-ranging power outage*. OFFICE OF TECHNOLOGY ASSESSMENT AT THE GERMAN BUNDESTAG. Retrieved May 12, 2021, from https://www.tab-beim-bundestag.de/en/pdf/publications/books/petermann-etal-2011-141.pdf

Pliatsios, D., Sarigiannidis, P., & Lagkas, T. (2020). A Survey on SCADA Systems: Secure Protocols, Incidents, Threats and Tactics. *IEEE*. Retrieved November 27, 2020, from https://ieeexplore.ieee.org/document/9066892

Pry, P. V. (2020, June 10). *China: EMP Threat: The People's Republic of China Military Doctrine, Plans, and Capabilities for Electromagnetic Pulse (EMP) Attack*. EMP Task Force on National; Homeland Security. Retrieved May 28, 2021, from https://apps.dtic.mil/sti/citations/AD1102202

Pry, P. V. (2021, January 28). *Russia: EMP Threat. The Russian Federation's Military Doctrine, Plans, and Capabilities for Electromagnetic Pulse (EMP) Attack*. DHS. Retrieved May 28, 2021, from https://apps.dtic.mil/sti/citations/AD1124730

Rahimi, K., & Davoudi, M. (2018). Electric vehicles for improving resilience of distribution systems. *Sustainable Cities and Society*, *36*, 246–256. https://doi.org/10.1016/j.scs.2017.10.006

Rahmandad, H., & Sterman, J. D. (2012). Reporting guidelines for simulation-based research in social sciences: Reporting Guidelines for Simulation-Based Research. *System Dynamics Review*, *28*(4), 396–411. https://doi.org/10.1002/sdr.1481

Reed, D. A., Powell, M. D., & Westerman, J. M. (2010). Energy Supply System Performance for Hurricane Katrina. *Journal of Energy Engineering*, *136*(4), 95–102. https://doi.org/10.1061/(ASCE)EY.1943-7897.0000028

Rinaldi, S. M. (2004). Modeling and Simulating Critical Infrastructures and Their Interdependencies. *th Hawaii International Conference on System Sciences*, 8.

Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE control systems magazine*, *21*(6). Retrieved November 25, 2020, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.2276&rep=rep1&type=pdf

Román, M. O., Stokes, E. C., Shrestha, R., Wang, Z., Schultz, L., Carlo, E. A. S., Sun, Q., Bell, J., Molthan, A., Kalb, V., Ji, C., Seto, K. C., McClain, S. N., & Enenkel, M. (2019). Satellite-based assessment of electricity restoration efforts in Puerto Rico after Hurricane Maria. *PLOS ONE*, *14*(6), e0218883. https://doi.org/10.1371/journal.pone.0218883

Sakurai, M., Watson, R. T., Abraham, C., & Kokuryo, J. (2014). Sustaining life during the early stages of disaster relief with a frugal information system: Learning from the great east Japan earthquake. *IEEE Communications Magazine*, *52*(1), 176–185. https://doi.org/10.1109/MCOM.2014.6710081

Sandberg, J., Holmström, J., & Lyytinen, K. (2020). Digitization and Phase Transitions in Platform Organizing Logics: Evidence from the Process Automation Industry. *Management Information Systems Quarterly*, *44*(1), 129–153.

Sanger, D. E., & Perlroth, N. (2021). Pipeline Attack Yields Urgent Lessons About U.S. Cybersecurity [newspaper]. *The New York Times*. Retrieved May 28, 2021, from https://www.nytimes.com/2021/05/14/us/politics/pipeline-hack.html

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the Is Discipline: Its Historical Legacy and Its Continued Relevance. *MIS Quarterly*, *43*(3), 695–A5. https://doi.org/10.25300/MISQ/2019/13747

Satariano, A. (2021). What is Fastly, the company behind the worldwide internet outage? [newspaper]. *The New York Times*. Retrieved June 15, 2021, from https://www.nytimes.com/2021/06/08/business/fastly-internet-outage.html

Savage, E., Gilbert, J., & Radaskey, W. (2010, January). *The Early-Time (E1) High-Altitude Electromagnetic Pulse (HEMP) and Its Impact on the U.S. Power Grid*. Metatech Corporation. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.734.5078&rep=rep1&type=pdf

Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., van de Koppel, J., van de Leemput, I. A., Levin, S. A., van Nes, E. H., Pascual, M., & Vandermeer, J. (2012). Anticipating critical transitions. *Science (New York, N.Y.)*, *338*(6105), 344–348. https://doi.org/10.1126/science.1225244

Seger, E., Avin, S., Pearson, G., Briers, M., Ó Heigeartaigh, S., & Bacon, H. (2020). *Tackling threats to informed decision-making in democratic societies – Promoting epistemic security in a technologically-advanced world*. Alan Turing Institute. Retrieved November 4, 2020, from https://www.turing.ac.uk/sites/default/files/2020-10/epistemic-security-report_final.pdf

Siegel, N. (2018). Using Systems Engineering to Create a Survivable Communications System that will Operate in the Presence of "Black Sky" Hazards. In A. M. Madni, B. Boehm, R. G. Ghanem, D. Erwin, & M. J. Wheaton (Eds.), *Disciplinary convergence in systems engineering research* (pp. 959–972). Springer International Publishing. https://doi.org/10.1007/978-3-319-62217-0_67

Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, *106*(6), 467–482.

Soltan, S., & Mittal, P. (2018). Blackiot: IoT Botnet of High Wattage Devices Can Disrupt the Power Grid. *Proceedings of the 27th USENIX Security Symposium*, 19.

Stockton, P., & Council, E. (2016). *Electric Infrastructure Protection (E-PRO®) Handbook* (Vol. 1). Electric Infrastructure Security (EIS) Council.

Stockton, P., & Council, E. (2018). *Electric Infrastructure Protection (E-PRO®) Handbook III* (Vol. 3). Electric Infrastructure Security (EIS) Council.

Taleb, N. N. (2010). *The black swan: The impact of the highly improbable* (2nd ed.). Random House Trade Paperbacks.

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder*. Random House.

Tarafdar, M., Cooper, C. L., & Stich, J.-F. (2019). The technostress trifecta - techno eustress, techno distress and design: Theoretical directions and an agenda for research. *Information Systems Journal*, *29*(1), 6–42. https://doi.org/10.1111/isj.12169

Tarafdar, M., Gupta, A., & Turel, O. (2013). The dark side of information technology use. *Information Systems Journal*, *23*(3), 269–275. https://doi.org/10.1111/isj.12015

Tehrani, P. M., Manap, N. A., & Taji, H. (2013). Cyber terrorism challenges: The need for a global response to a multi-jurisdictional crime. *Computer Law & Security Review*, *29*(3), 207–215. https://doi.org/10.1016/j.clsr.2013.03.011

Traywick, C., Chediak, M., Malik, N. S., & Saul, J. (2021). The Two Hours That Nearly Destroyed Texas's Electric Grid [newspaper]. *Bloomberg.com*. Retrieved May 12, 2021, from https://www.bloomberg.com/news/features/2021-02-20/texas-blackout-how-the-electrical-grid-failed

U.S. Department of Energy. (2014). *Large Power Transformers and the U.S. Electric Grid (2014 Update)*. Retrieved May 9, 2022, from https://www.energy.gov/sites/default/files/2014/04/f15/LPTStudyUpdate-040914.pdf

Ustun, T. S., Cali, U., & Kisacikoglu, M. C. (2015). Energizing microgrids with electric vehicles during emergencies — Natural disasters, sabotage and warfare. *2015 IEEE International Telecommunications Energy Conference (INTELEC)*, 1–6. https://doi.org/10.1109/INTLEC.2015.7572377

Vaijhala, S., & Rhodes, J. (2018). Resilience Bonds: a business-model for resilient infrastructure. *Field Actions Science Reports. The journal of field actions*, (Special Issue 18), 58–63.

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Wang, W., & Lu, Z. (2013). Cyber security in the Smart Grid: Survey and challenges. *Computer Networks*, *57*(5), 1344–1371. https://doi.org/10.1016/j.comnet.2012.12.017

Wessel, L., Baiyere, A., Ologeanu-Taddei, R., Cha, J., & Jensen, T. B. (2021). Unpacking the Difference Between Digital Transformation and IT-Enabled Organizational Transformation. *Journal of the Association for Information Systems*, *22*(1). https://doi.org/10.17705/1jais.00655

Wilkinson, A., Kupers, R., & Mangalagiu, D. (2013). How plausibility-based scenario practices are grappling with complexity to appreciate and address 21st century challenges. *Technological Forecasting and Social Change*, *80*(4), 699–710. https://doi.org/10.1016/j.techfore.2012.10.031

Wilson, C. (2008, July 21). *High Altitude Electromagnetic Pulse (HEMP) and High Power Microwave (HPM) Devices: Threat Assessments*.

Wolff, J. (2020). Opinion — How Ransomware Puts Your Hospital at Risk [newspaper]. *The New York Times*. Retrieved May 28, 2021, from https://www.nytimes.com/2020/10/17/opinion/hospital-internet-security-ransomware.html

Wright, A. (2005). The role of scenarios as prospective sensemaking devices (K. Chaharbaghi, Ed.). *Management Decision*, *43*(1), 86–101. https://doi.org/10.1108/00251740510572506

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting*, *25*(4), 813–825. https://doi.org/10.1016/j.ijforecast.2009.05.019

Yoo, Y. (2010). Computing in Everyday Life: A Call for Research on Experiential Computing. *MIS Quarterly*, *34*(2), 213. https://doi.org/10.2307/20721425

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research commentary—the new organizing logic of digital innovation: An agenda for information systems research. *Information systems research*, *21*(4), 724–735.

Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of Things for Smart Cities. *IEEE Internet of Things Journal*, *1*(1), 22–32. https://doi.org/10.1109/JIOT.2014.2306328

Zeiss, R., Ixmeier, A., Recker, J., & Kranz, J. (2021). Mobilising information systems scholarship for a circular economy: Review, synthesis, and directions for future research. *Information Systems Journal*, *31*(1), 148–183. https://doi.org/10.1111/isj.12305

Zuboff, S. (2019, January 31). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.

Appendix

Affidavit

according to § 8 paragraph 3 of the doctoral degree regulations of

February 17, 2015

"I hereby affirm in lieu of oath that I have prepared the submitted work independently and without using any other than the specified aids. Statements, data and concepts taken directly or indirectly from other sources are identified with reference to the source. In the selection and evaluation of the following material, the persons listed below helped me in the manner described, for a fee / free of charge (underline as appropriate):

Other people, in addition to the co-authors listed in the introduction to the work, were not involved in the content-related preparation of the present work. In particular, I did not make use of the paid help from mediation or advisory services. Nobody has received direct or indirect monetary benefits from me for work that is related to the content of the submitted dissertation.

The thesis has not yet been submitted to another examination authority in the same or a similar form, either in Germany or abroad.

I assure you that to the best of my knowledge, I have told the pure truth and have not concealed anything.

I confirm that the submitted electronic version corresponds in full to the submitted printed version.

I am aware of the criminal liability of a false affidavit, namely the threat of punishment according to § 156 StGB with an imprisonment of up to three years or with a fine in the case of intentional commission of the act or according to § 161 Paragraph 1 StGB up to one year imprisonment or fine in the case of negligent commission.

Cologne, 15.01.2024

Place, Date   Signature