Understanding the Digital Transformation. Philosophical Perspectives on Potentially Gradual Disruptions

Armin Grunwald

Abstract: This paper introduces the phenomenon of possible 'gradual disruptions' in quickly ongoing digitalization. This type of disruption does, in contrast to the familiar meaning of 'disruption', not occur unexpectedly and suddenly, but builds up gradually over longer timespans. Finally only, the sudden disruption of familiar constellations may lead to damage. The digital transformation, so my thesis, increases both the possibility and the risk of gradual disruption. As an example, the dependency of society and humankind in total on functioning digital technology infrastructures steadily increases. Hence, societies become gradually more vulnerable. Total dependencies are latent disruptions: high damage will occur in the case of the breakdown of the infrastructures. Gradual disruptions, however, are difficult to detect because of their slow emergence. Philosophy can contribute to early recognition in epistemic, ethical, pragmatic, and communicative dimensions.

1. Digitalization as a narrative and process

1   Modern societies have been experiencing fast change according to science-based and technological innovation. This phenomenon which well describes what many people perceive is not quite new as the following quote demonstrates:

2   Domestic life, political institutions, international relations and personal contacts are shifting with kaleidoscopic rapidity before our eyes. We cannot appreciate and weigh the changes; they occur too swiftly. We do not have time to take them in. No sooner we begin to understand the meaning of one such change than another comes and displaces the former. (Dewey 1931, 53)

3   Since decades, this attitude to life has been strengthened by ongoing and accelerating digitalization. Thinking digitality, therefore, also includes understanding modern society in this contemporary situation.

4   The term digitalization serves different purposes on two levels. On the one hand, it is about describing empirically observable phenomena, such as the use of digital tools in the areas of work, leisure, and public communication as well as the consequences of this use. On the other hand, digitalization represents a dominant *narrative of* social debates with partly visionary, partly fatalistic, and partly normative intentions. This is expressed, for example, in the diagnosis of a "digital determinism" (Mainzer 2016). Belief in this determinism motivates politicians and businesspeople postulating that society must make itself 'fit for digitalization'. Another expression of this narrative are trans- and post-humanist ideas of overcoming deficient human civilization through a supposedly perfect technological civilization based on accelerated digitalization (Hurlbut and Tirosh-Samuelson 2016).

5   This paper focuses on specific aspects in exploring the philosophical view of digitalization as a *social process* (Grunwald 2019). Digitalization as a process refers to the digitally enabled transformation of social contexts, for example, of public communication and democracy, of the transformation of the economy to a data- and knowledge-based economy, of new and emerging human-machine relationships, of changes in the world of labor, and of human self-perceptions affected by digital technologies and Artificial Intelligence (AI).

6   Facing and experiencing rapid digitalization, hopes and concerns about its possibly disruptive character repeatedly have become the subject of scientific, philosophical, and public debates. This paper deals with – this formulation may sound paradoxically – *gradual disruptions* at the societal level. The notion of gradual disruption refers to upheavals with considerable or even dramatic damage potential that do not occur unexpectedly and suddenly like a global pandemic or a war of aggression, but rather build up gradually until they finally lead to the disruption of formerly stable constellations (Chap. 2). It will be argued that this type of potential and gradual disruption can be assumed to take place in many areas of digitalization (Chap. 3). Finally, proposals for the corresponding research agenda for a philosophy of digitalization will be presented (Chap. 4).[1]

---

1. This paper develops previous work of the author further and is based, in particular, on (Grunwald 2019) and (Grunwald 2023), both in German language.

2. Disruption as an analytical concept

7   The career of the disruption concept started with the theory of disruptive technology (Bower and Christensen 1995). It was quickly extended to the field of innovation (e.g., (Danneels 2004)) by talking about breakthroughs and disruptive innovation. The definition as well as assumptions and expectations are contested (Gans 2017): "'Disruption' is a business buzzword that has gotten out of control. Today everything and everyone seem to be characterized as disruptive – or, if they aren't disruptive yet, it's only a matter of time before they become so" (MIT Press 2017). In this critique, the concept of disruption is flattened into a synonym for success which, however, has not changed the successful career of the concept to date. In particular, disruption has become a widely used term in the last ten years at the occasion of accelerated digitalization and new business models. Although the word's origin refers to rather unpleasant-sounding meanings (lt. *disrumpere* = to burst, break, tear apart), it was initially used mostly with positive intentions. Still, disruptive innovations as technological leaps or paradigmatic changes in business models are high on the agenda of innovation policy. In Germany, even an own institution, the Federal Agency for Disruptive Innovation (SPRIND), was founded in 2019 by the national government. It shall support fast and expectedly disruptive innovation to strengthen Germany's national economy.

8   In sharp contrast to the appreciation of disruption in innovation policy, crisis phenomena have been placed in the context of disruption more recently. Above all, the coronavirus pandemic and the Ukraine war are considered disruptive events. Both have put an end to a long period of far-reaching stability, at least in the Global North. They indicate, according to widespread diagnosis, the transition to a period of permanent crises, associated e.g. with climate change, migration, and new geopolitical conflicts. In this way, the term disruption is used to describe the breakup of formerly stable social conditions with the often-feared consequence of dramatic, negative effects. Indicative of this is the currently widespread communication of catastrophic narratives in public debate.

9   In both meanings mentioned, disruption refers to the sudden breakup of well-known constellations with formerly long-term stability. Expectations of stability, assumptions of continuity, and planning security are shattered, casting the prospects for the future in an uncertain light. The bursting, breaking, and tearing rooting in the Latin word (see above) semantically refers to the time structure of more or less sudden, abrupt events. Seen in this light, the talk of *gradual* disruption in the title of this papter seems conceptually wrong, contradictory, or paradoxical.

10   A closer look allows us to differentiate. Semantically, the meaning of the term "disruption" consists of two parts: on the one hand, the *breaking up of* previously stable conditions and, on the other hand, the *speed* of this breaking up. While the first component is, following its Latin ori-

gin, etymologically inherent to the term, the second can be handled more flexibly. Time scales of breaking can be elastic. For example, the invention of the printing press in the late Middle Ages is often portrayed as disruptive one – historically, however, this disruption extended over many decades of diffusion into the societies of the time. Disruption and processes of breaking up can also build up *gradually* over longer periods of time and only in the course of time lead to developments which are *ex post* regarded as seemingly sudden disruptions in the sense of a qualitative break.

11   The Russian attack on Ukraine in 2022 offers a further view on disruption. After a sudden break, people sometimes argue that there had been some indications of the coming disruption in advance, weak signal perhaps which were overseen or ignored. After the start of the Ukraine war, it was widely noted that it did not happen without warning, but had a history in the form of Russian attacks on territories of the former Soviet Union now belonging to independent states. Many examples of disruption heralded by recognizable, but often unrecognized, signs are known from the technical world, especially material fatigue and wear. The daily stress on many technical objects, such as V-belts in older vehicles or bridge structures, gradually leads to wear and degradation. Nevertheless, they still function reliably for a long time until the wear reaches a point where the component fails from one moment to the next, i.e., in the chosen examples the V-belt breaks or the bridge collapses. An example from the climate debate are the so-called *tipping points* (Gladwell 2000). With further warming, self-reinforcing feedback effects could appear, which would have dramatic consequences within a short period of time, i.e., a disruptive effect. The disruptive in events of this type is thus inherent in incremental processes that are difficult to detect. Therefore, they escape early intervention and preventive action, can slowly evolve further and can suddenly lead to potentially far-reaching consequences. The tragedy of such gradual developments is, one could say with slightly existentialist emphasis, that in the incremental course, serious disruptions can announce themselves insidiously but then take place abruptly.

12   Using these semantic differentiations, we will consider examples for *possibly gradual disruption* in the digital transformation. The focus will be on possible developments with damaging or even catastrophic potential, not on the question of desired (at least by some actors) disruptive innovation. The reflections do not consider the probability of the emergence of such disruptions but remain in the field of their possibility.

3. Disruptive potentials of the digital transformation

13   The sometimes rapid success of many developments in the context of digitalization, for example, the global expansion of social media within a few years around 2010 or the current diffusion of AI applications, leads to shifts in many areas. These include, for example, human-technology

relations, distribution of responsibilities, industrial production, security and surveillance, public communication and political opinion formation, solidarity and competition, the labor market as well as leisure and media consumption. Four examples of such potential shifts will be discussed with respect to potential and gradual disruption: the standardization of human action (Sect. 3.1), increasing dependency (Sect. 3.2), the loss of the future (Sect. 3.3), and the possible end of opportunities for learning and reflection (Sect. 3.4).

3.1 Standardization of human action

14   In the common understanding, technology and the services based on it are supposed to serve as a means to human ends, satisfy needs, and solve problems in order – according to the narrative since the European Enlightenment – to expand the range of options for human action and promote human emancipation. However, this is only half the truth. For while technology expands human options for action, it simultaneously leads to a wide range of adaptation needs, up to coercion (Grunwald 2019b).

15   Technical systems structure and regulate human actions, for example, through operating instructions, regulations, and user interfaces. In many fields, this is trivial, for example, when certain physical movements are required to use a spade for digging or when the operating instructions must be followed to operate a washing machine. Digital technology, however, subtly changes human actions and behavior, possibly without being noticed. The debate around "software as institution" (Orwat, Raabe, and Buchmann 2010, 201) has drawn attention to the fact that software systems can have regulatory power, e.g., by regulating transactions or access and usage rights. For example, privately run social media platforms structure public communication, search engines use algorithms developed by private companies to sort their users' perception of the world, and online platforms structure business processes and crowd sourcing.

16   Specific adaptation needs arise in the interaction of autonomous software and robotic systems with humans. In the Industry 4.0 world, in which robots are expected to work together with humans in industrial production, the required communication between humans and robots must be free from misunderstandings for reasons of functionality and safety. This holds similarly for autonomous driving in mixed traffic with human road users. The requirement

17   To enable efficient, reliable, and safe communication between humans and machines and to avoid excessive demands, systems must adapt more closely to the communication behavior of humans and not, conversely, demand increased adaptations of humans (Ethikkommission 2017, 13).

18   is easy to raise and anthropologically understandable (Deutscher Ethikrat 2023). However, there are concerns that this will not happen in the real world, but that progressive digitalization will gradually regulate and standardize human actions according to the requirements of technical systems and technical communication. Despite all efforts to ensure human autonomy and to preserve freedom of choice and self-determination, the opposite could happen, namely a creep-

7

ing and unnoticed loss of freedoms. The maelstrom of safety thinking is, in particular, cited as evidence for this assumption, as in the case of autonomous driving. In the context of slippery slope arguments, the postulate of safety could ultimately completely undermine human freedom as result of the technical implementation of safety Similarly, surveillance in the private and public spheres, enabled and enforced by digital technology, is repeatedly justified with security arguments over freedom arguments.

19   The gradual disruption in these fields would be a slow and unnoticed slide into a world in which the security interests of the state move to the top of the hierarchy of aspects to weigh up, resulting in ever more digitally enforced standardizations of human action, which could ultimately lead to a shift away from freedom-oriented individualism to a controlled collectivism.

3.2 Dependence as latent disruption

20   Modern societies are already completely dependent on the smooth functioning of critical infrastructures such as electric power supply (Petermann et al. 2011). Increasingly, this also applies to digital infrastructures. If the internet failed, financial transactions would become impossible, the global economy would collapse, media communication would no longer be possible, medical diagnostics would be deprived of many established procedures, international logistics chains would come to a standstill, and much more. The increasing introduction of automated decision-making (ADM) systems creates a dependence on AI-controlled systems, which, together with their black box character and lack of transparency, but also due to the psychological *automation bias* (Safdar, Banja, and Meltzer 2020), (Deutscher Ethikrat 2023), means an increasing dependence on these systems in decision-relevant contexts such as the police and social services.

21   The gradual displacement of cash is a current example of the ambivalence of technical infrastructures. While cashless payment transactions were initially an *additional* option besides cash as a convenience for businesses and private individuals, there is now a gradual transition to a world without cash. Once cashless payment transactions have become fully established, as is already largely the case in some countries, this payment option may be the only one left, without any choice. If the internet would collapse, purchases or payment transactions would no longer be possible. While cashless payment was initially an additional option and increased choice, it gradually became dominant, finally it may become the sole option due to the disappearance of cash and thus a compulsion, accompanied by full dependence on the smooth functioning of all the technologies in the background.

22   Dependencies are not disruptions in themselves, but they carry their seeds. Dependencies that have become total are *latent disruptions*. As disruptions on demand, they build up gradually through increasing dependencies, but in an emergency, i.e., if, in this case, digital technologies were no longer to function smoothly, they could have abrupt and possibly catastrophic consequences. Obsering this possibility, naively relying on their unlimited smooth functioning and making the functionality and stability of modern societies dependent on this is a bet "for the whole" in the sense of Hans Jonas (Jonas 1979). Correspondingly, this approach is highly prob-

lematic with respect to responsibility. Unexpected hacking events, a collapse of state order, or severe economic turmoil could also affect critical infrastructures such as the internet and, in the worst case, push them into dysfunctionality. Even though it is usually difficult, if not impossible, to identify the point in the gradual development at which complete dependency sets in, this point has probably long since been passed with regard to many digital infrastructures and platforms – which means that modern societies are already in a state of latent disruption.

3.3 Loss of the future

23 Digital technologies often are regarded synonymous with the future. However, digital technologies generally operate based on past data. Thus, the digital twins of real-world objects can only mirror a world of yesterday. E.g., customer profiles can be created exclusively based on past purchase and consumption processes. Also, Big Data mining can only analyze past data and recognize past patterns. AI systems have to be trained on data from the past, as data from the future is not available. Due to the indispensable reference to data, digital technology is inescapably bound to past conditions. When data sets, digital twins, and correlations and patterns detected by AI are used to predict future developments, past facts and patterns are carried over into the future, are imposed on it. The future as an at least partially open space of alternative paths and possibilities is replaced by a data-based extension of the past (Grunwald 2023).

24 Digitalization or some of its fields could become conservative in this way, aligning future prospects with old data rather than developing new ideas. Multiple anthropological determinations understand humans as beings with a future and the ability to envision possible futures (e.g., (Kamlah 1973)), which go beyond extending the past to the future but, rather, include creative ideas in an open space of possibilities. They even may have a visionary, counterfactual and utopian character. A gradual disruption could occur here if the fundamental openness of the future fades into the background or disappears completely in favor of a data-driven orientation bound to the past.

3.4 End of reflection and learning opportunities

25 Acceleration is part of the capitalist economic system. It unleashes creativity and innovation, especially through competition. Acceleration is a phenomenon often discussed in the context of digitalization, but was already an issue a hundred years ago (see the quote at the top of this paper). The increase in computing speed, the possibility of calculating millions of options in the shortest time, the linking of creative resources via the internet, and the acceleration of data transfer and communication, much of it mediated and further accelerated by means of digital twins, shorten innovation cycles. For some years now, as briefly mentioned (Chap. 2), disruptive innovation has therefore held fascination as extreme acceleration which is the opposite of incremental innovation processes.

26 One concern regarding digitalization relates to negative and potentially ruinous consequences of

ever-increasing acceleration, in particular to the question of whether and when further acceleration could fundamentally undermine the conditions of reflection. Reflexivity needs anticipatory and precautionary engagement with consequences for the choice of actions and decisions at a wide variety of levels. It implies the anticipation of their consequences before action is taken, as well as the implementation of its results in the subsequent decision-making processes. It requires careful analysis and deliberation, the transparent weighing of alternatives, the search for the right balance and ethically legitimate criteria for choice-making. All of this takes time in two ways: first, for the deliberation and consideration processes themselves and, second, for transferring the results into practical action and decision-making, e.g. by scientific policy advice (Grunwald 2019b). Otherwise, the human being could no longer be a *zoon politikon* and a moral being with reflection, deliberation, and dialogue (Deutscher Ethikrat 2023).

27   The gradual disruption in this respect could be that societal capacities and structures for reflection would slowly erode with the argument of having no time because of capitalist competition. In the narrative of an innovation-oriented fatalism under the primacy of competitive thinking, reflection could no longer be afforded, since otherwise the competitor would be faster and gain market advantages.

4. Philosophical inquiries into potentially disruptive developments

28   The above examples provide information about the phenomenology of potentially gradual disruptions or gradual developments with disruptive potential in the digital transformation. Overarching philosophical patterns emerge, in particular, with respect to (1) epistemology, (2) assessment, and (3) pragmatics.

29   *(1) Epistemological dimension: How to recognize gradual disruption?*

30   Part of the phenomenology of gradual, creeping developments is that they are usually difficult to detect. This is a challenge, especially in their early phases, because then the availability of data is limited and at best reveals weak signals. Merely possible developments with disruption potential interact in an unknown future with other merely possible developments in equally unknown ways, creating an epistemological situation of high complexity and uncertainty. Due to the weak evidence of these data and a lack of sensitivity to the only slowly emerging disruption potential, it can be difficult to encourage systematic research to clarify the facts and to mobilize the required funding. Many aspects remain rather speculative at that stage. Only as the development goes on and comes closer to the possible occurrence of a disruption, the evidence increases, the effects become more visible and the scientific understanding of the interrelationships becomes more profound. With regard to the state of knowledge on climate change as such a gradual disruption, this has been observed repeatedly over the last forty years. The epistemological chal-

10

lenge is to improve the opportunities to extract reliable knowledge out of a weak and deficient data basis, e.g. by an epistemology of correlation in view of using digital means of modeling and data mining to detect weak signals of disruptive developments as early as possible.

31 *(2) Ethical dimension: How to classify and assess?*

32 The epistemological constellation has direct consequences for the evaluation and classification of the development that only gradually becomes visible. The seemingly obvious conclusion that ethics should hold back until better knowledge is available (Nordmann 2007) is no solution given the high relevance of possible disruption. In view of limited resources, various slow-moving developments and feared events must be compared and ranked according to urgency. Prioritizations and urgency assessments, which are based on normative criteria and relevance, e.g., with regard to the *Leitbild* of sustainable development, are, however, related to the evidence of knowledge. A mere suspicion is not sufficient for a high prioritization with, e.g., the resulting allocation of resources, even if the suspicion would address a development that would be disastrous according to recognized standards and would definitely have to be averted. Here, difficult necessities of assessing the situation and classifying it in comparison with other developments arise and need normative reflection.

33 *(3) Pragmatic dimension: How to act?*

34 In early phases of possible disruption, questions arise about consequences for action between proactively intervening prevention and waiting for better data and clearer diagnoses. Here, the Collingridge dilemma (Collingridge 1980) familiar in technology assessment (Grunwald 2019b) must be taken into account: In principle, the further course of events is still wide open in early phases and can therefore be influenced more easily than in later phases, when the constellation is already strongly entrenched by path dependencies. However, the required knowledge of the consequences of developments suspected of gradual disruption is then inevitably highly uncertain or completely lacking. Instead of reliable forecasts or at least plausible scenarios, there are usually only more or less speculative expectations or even fears whose epistemic content can often hardly be assessed (Grunwald 2013). In this respect, gradual disruptions present themselves as a radicalization or even an extreme form of the Collingridge dilemma: Before the disruptive event, little or nothing is known about the consequences, so that no preventive or proactive actions can be initiated; afterwards, however, the event has happened and it is too late for preventive measures. The requirement of reflexivity (Sect. 3) would thus be violated. In view of this situation, the question arises as to how much evidence of a suspicion is needed to legitimize intervening measures, to mobilize budgets, to limit freedoms through regulation, if necessary, etc.

5. Epilogue

35 The possibly gradual disruptions outlined briefly above may seem to be a nightmare as the dark

side of the digital transformation. However, they are no more than *possibilities* based on current observations and concerns. Reflecting about such possibilities follows the purpose of creating awareness and enabling the view on possible dark sides of the digitalization in contrast to all the powerful positive narratives (cp. Sect. 1). The aim is not to downplay the many positive consequences of the digital transformation but, rather, to support gaining the full picture which is needed to shape reflectively and responsibly its future path. Philosophy offers significant contributions to this reflected view.

Bibliography

Bower, Joseph L., and Clayton M. Christensen. 1995. "Disruptive Technologies. Catching the Wave." *Harvard Business Review* 96:19–45.

Collingridge, David. 1980. *The Social Control of Technology*. New York: Pinter.

Danneels, Erwin. 2004. "Disruptive Technology Reconsidered. A Critique and Research Agenda." *Journal of Product Innovation Management* 21 (4): 246–58.

Deutscher Ethikrat. 2023. "Mensch Und Maschine – Herausforderungen Durch Künstliche Intelligenz." https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf.

Dewey, John. 1931. "Science and Society." In *The Later Works 1925-1953*, 6:53–63. Carbondale: Southern Illinois University Press.

Ethikkommission. 2017. "Autonomes Und Vernetztes Fahren. Bericht." https://bmdv.bund.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile.

Gans, Joshua. 2017. *The Disruption Dilemma*. Cambridge: MIT Press.

Gladwell, Malcolm. 2000. *The Tipping Point – How Little Things Can Make A Big Difference*. New York: Little, Brown and Company.

Grunwald, Armin. 2013. "Modes of Orientation Provided by Futures Studies: Making Sense of Diversity and Divergence." *European Journal of Futures Research* 2 (30): 1–9. https://doi.org/10.1007/s40309-013-0030-5.

———. 2019a. "Digitalisierung Als Prozess. Ethische Herausforderungen Inmitten Allmählicher Verschiebungen Zwischen Mensch, Technik Und Gesellschaft." *Zeitschrift Für Wirtschafts- Und Unternehmensethik* 20 (2): 121–45. https://doi.org/10.5771/1439-880X-2019-2-121.

———. 2019b. *Technology Assessment in Practice and Theory*. Abingdon: Routledge.

———. 2023. "Digitalisierung Als Prozess: Der Philosophische Blick Auf Die Möglichkeit Allmählicher Disruption." In *Philosophie Der Digitalisierung. Eine Forschungsagenda*, edited by Matthias Kettner.

Hurlbut, J. Benjamin, and Hava Tirosh-Samuelson, eds. 2016. *Perfecting Human Futures. Transhuman Visions and Technological Imaginations*. Wiesbaden: Springer.

Jonas, Hans. 1979. *Das Prinzip Verantwortung. Versuch Einer Ethik Für Die Technologische Zivilisation*. Frankfurt/M.: Suhrkamp.

Kamlah, Wilhelm. 1973. *Philosophische Anthropologie*. Braunschweig: Bibliographisches Institut.

Mainzer, Wilhelm. 2016. *Wann Übernehmen Die Maschinen*. Heidelberg: Springer.

MIT Press. 2017. "Announcement of the Book The Innovation Dilemma by Joshua Gans." https://mitpress.mit.edu/9780262533621/the-disruption-dilemma/.

Nordmann, Alfred. 2007. "If and Then: A Critique of Speculative NanoEthics." *NanoEthics* 1:31–46.

Orwat, Carsten, Oliver Raabe, and Erik Buchmann. 2010. "Software Als Institution Und Ihre Gestaltbarkeit." *Informatik-Spektrum* 33:626–33.

Petermann, Thomas, Harald Bradke, Arne Lüllmann, Maik Poetzsch, and Ulrich Riehm. 2011. *What Happens during a Blackout. Consequences of a Prolonged and Wide-Ranging Power Outage*. Technology Assessment Studies Series 4. Norderstedt: BoD – Books on Demand.

Safdar, Nabile M., John D. Banja, and Carolyn C. Meltzer. 2020. "Ethical Considerations in Artificial Intelligence." *European Journal of Radiology* 122:108768.