Socially Disruptive Technologies, Moral Progress, and Rule Following

Lane, Benedict

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

RESEARCH ARTICLE

Socially Disruptive Technologies, Moral Progress, and Rule Following

Benedict Lane[1]

Abstract

One of the most ethically significant features of new and emerging 21st century technologies is their potential to disrupt the social status quo, for better or worse. Correspondingly, one of the most pressing questions in the philosophy and ethics of technology is how to understand and respond to this potential for social disruption. A prominent account of social disruption is what I call the "epistemic account," according to which socially disruptive technologies are *sources of moral-epistemic uncertainty* and *trigger moral inquiry* into the novel circumstances they bring about and the new actions they make possible. This paper begins with a critique of the epistemic account and develops an alternative account according to which socially disruptive technologies undermine the *very possibility of progressive inquiry* (as opposed to simply triggering it). It continues to draw attention to an underappreciated connection between the concept of social disruption, the notion of progress through inquiry, and our capacity to "go on in the same way," drawing on the literature surrounding Wittgenstein's famous rule-following considerations. A picture of social disruption is sketched out – the "language-game" picture – that challenges many widespread philosophical assumptions, including the objectivity of disruptive power, the explainability and predictability of social disruption, and the justificatory status of our ethical responses to social disruption.

Keywords Socially disruptive technologies · Conceptual disruption · Moral progress · Rule following · Ludwig Wittgenstein

✉ Benedict Lane
   j.b.h.lane@tudelft.nl

[1] Faculty of Technology, Policy & Management, TU Delft, Delft, Netherlands

 Springer

1 Introduction

Technology both shapes and is a component of human beings' contemporary form of life (Winner, 2014), acting as an intermediary between us and our world (Ihde, 1990). When technology changes– whether incrementally or radically– our form of life therefore changes with it. Sometimes such technology-induced changes are easy to cope with, and other times they aren't; when a technology induces changes to our form of life that are not easy to cope with, we call it socially disruptive (Hopster, 2021). Technology is not the only source of social disruption– Hopster et al. (2023) mention "intercultural dialogue and interaction" as just one other non-technological source. However, the rapid acceleration of technological advancement, and the proliferation of new and emerging science and technology (Swierstra & Rip, 2007), makes socially disruptive technologies (SDTs) arguably the dominant source of social disruption in the contemporary world, and therefore a fitting focal point for philosophical discussion of the phenomenon of social disruption generally.

While often acting as a means to achieving various moral goals, technological change can affect our moral practices themselves, bringing certain features of the world into the moral foreground and inviting certain courses of action in response to perceived moral situations (Verbeek, 2006, 2008). This results in a process of "techno-moral change" (Danaher & Sætra, 2023), affecting many important aspects of moral life, including, for example: the practical options we perceive as open or closed to us, or even demanded of us (Latour, 1992), and the inequalities thereof or discrimination therein (Winner, 1980); how and with whom– or even with *what* (Danaher, 2019; Sætra, 2022)– we can form meaningful relationships (Cocking & Matthews, 2000; Briggle, 2008), and what exactly we owe to those into whom we enter into such newly possible relationships (Danaher, 2017; Vallor, 2012); and how we understand what it means for us to live well in a world that is increasingly technologically constituted (Vallor, 2016). Given its evident fundamentality, such techno-moral change also has the potential to be socially disruptive.

In this paper, I develop an original account of social disruption, with a focus on techno-moral disruption. I motivate my account through critical engagement with the existing literature on social disruption and SDTs, arguing that the current best philosophical understandings of social disruption and SDTs are either inadequate or incomplete. Having assessed the merits of existing accounts of social disruption and SDTs, I develop a new account: I argue that social disruption is an interruption to the possibility of effective collective inquiry, and I connect the possibility of such inquiry to a particular understanding of what it means to follow a rule. Drawing heavily on collectivist interpretations of Wittgenstein's rule-following considerations, I propose that social disruption ought to be understood as the breakdown of what I call "deep agreement" among a community of inquiry. The underlying idea is that for societal inquiry to make progress, its participants "must already judge in conformity with mankind" (Wittgenstein, 2008, § 156). When such conformity in judgement is absent, progress is impossible; inquiry cannot move forward because it is marred in something like what Kuhn called "debate over fundamentals" (2012, p. 91). Understanding social disruption in this way forces us to quite radically reconsider the explanatory role we often assign to SDTs in bringing about their socially disruptive

effects, the objective light in which often cast the phenomenon of social disruption, as well as what it means to respond appropriately to an episode of social disruption.

The paper is structured as follows. In Sect. 2, I propose five desiderata for an analysis of social disruption that will both shape the discussion of the rest of the paper and, I hope, set the agenda for future discourse on social disruption and SDTs. In Sect. 3, I critically assess the current literature on social disruption and SDTs in terms of the proposed desiderata. I offer reasons to reject a prominent account of social disruption– the epistemic account. In Sect. 4, I argue in favour of a "deep" or pre-epistemic account, and I establish a connection between the possibility of progress through inquiry and the notion of rule-following, a connection that has so far only been hinted at in the literature. In Sect. 5, I develop an original picture of social disruption– what I call the "language-game" picture. I contrast the language game picture with a prima facie plausible alternative, what I call the "rules-as-rails" picture. I offer reasons for rejecting "rules-as-rails" in favour of the language-game picture on the basis of the aforementioned desiderata. However, I demonstrate that even the language-game picture cannot satisfy all of the proposed desiderata simultaneously, and I conclude that a proper understanding of rule-following in the context of social disruption and SDTs ought to temper our expectations of what an account of social disruption can do. I call for greater attention to be paid to the role of rule-following in future discourse around social disruption and SDTs.

2 Five Desiderata for an Analysis of Social Disruption

Before delving into the existing literature on social disruption and SDTs, I want to propose five desiderata in terms of which it seems reasonable to judge any account of social disruption. Desiderata are important for setting the agenda of a debate and for providing clear standards of philosophical success. Obviously, choosing the standards of success for a task one is setting oneself comes with the same risks as marking one's own homework. However, one of the most interesting things about the desiderata that I propose is that they are extremely difficult to mutually satisfy– indeed, the positive account of social disruption I develop in the second half of the paper fails to satisfy them all, prompting me to call for a reassessment of what we can reasonably expect an account of social disruption to deliver. This admission should, I hope, alleviate the suspicion that I am moving the goalposts in my own favour (for, surely, I would have done a better job of it if I were).

The desiderata in terms of which I suggest we ought to judge any account of social disruption are based on a loose and imperfect analogy with the medical concept of health and the philosophical debates surrounding it. The metaphor is, consistent with my overall approach, a Wittgensteinian one. As he puts it in an evocative passage from *Remarks on the Foundation of Mathematics*:

> The sickness of a time is cured by an alteration in the mode of life of human beings, and it was possible for the sickness of philosophical problems to get cured only through a changed mode of thought and of life, not through a medicine invented by an individual. (1996, II §23)

This pregnant metaphor serves as my starting point for thinking about social disruption. The desiderata I propose for judging an account of social disruption are as follows:

1. It would differentiate between a domain of inquiry that is functioning "healthily," in a way that is conducive to progress,[1] from a practice that is experiencing disruption.

When I suggest that a socially disrupted domain of inquiry is "unhealthy," what I mean is that disruption– analytically– is not part of the perceived "natural," "normal" flow of things; disruption is an experienced obstruction to forms of worthwhile activity.[2] What this means exactly is left deliberately open at this stage. Indeed, it has been a longstanding subject of philosophical debate whether and to what extent the medical concept of health is grounded in brute, biological reality or in a contingent, local system of social stances and attitudes, i.e., whether health is a natural or a social kind, such that what is regarded as a disease from one perspective might be regarded as "natural" and "normal" from another (e.g., Boorse, 1975, 1977; Margolis, 1976; Reznek, 1987; Cherry, 1996). The health metaphor is intended to leave open the analogous question of whether "healthy inquiry" is itself a natural or a social kind.

The health metaphor also draws attention the normative status of disruption as the *absence* of healthy inquiry; the issue of the normativity of the medical concept of health has been widely discussed in the philosophical literature (e.g., Goosens, 1980; Lennox, 1995). H. Tristram Englehardt claimed that "choosing to call a set of phenomena a disease involves a commitment to medical intervention" (1975, p. 137). A disease, on this view, is something to be prevented, cured, or treated. On an analogous view, choosing to call a set of phenomena an episode of social disruption arguably involves a commitment to preventing, eliminating, or resolving it. In this regard, social disruption is *pro tanto* bad, in the sense that it poses a problem to be overcome.[3] However, in consequentialist terms, social disruption can be good *in toto*: it can lead to a progressive change in a community's practices (in the same way that an episode of unhealth can prompt an individual to adopt a healthier lifestyle). I return to this point in desideratum 4.

---

[1] It has been suggested to me that a healthy domain of inquiry might also be one that is simply not exhibiting regress; stasis can be healthy too. I agree that stasis can sometimes be healthy. However, inquiry in healthy stasis must be actively kept "ticking over" and not passively left dormant. (A corpse, while not exhibiting regress, is not in good health.) Active stasis still requires continued participation in the relevant activities of inquiry; passive stasis does not. Thus, anticipating my argument below, the healthy functioning of a domain of inquiry– whether healthily static or progressing– should be understood in terms of the possibility of our continued participation in such activities.

[2] As an anonymous referee conscientiously pointed out to me, it is important for my metaphor not to overlook the reality of chronic illness. There is a genuine sense in which chronic illness becomes the "normal" state for someone experiencing it. However, a chronic illness still represents an experienced obstruction to certain forms of worthwhile activity that are available for others (otherwise, arguably, it would not be an illness but simply an alternative mode of being). It is an interesting question whether a form of life can experience chronic unhealth and what this would mean, in my metaphorical sense– but it is beyond the scope of this paper to stretch the metaphor so far.

[3] Thank you to an anonymous referee for making sure I got this aspect of the metaphor straight.

2. It would accurately capture the individual phenomenology of what it is like to be a participant in a domain of inquiry that is socially disrupted.

Sticking with our metaphor, it has been recognised by some philosophers that health and disease have an important phenomenological component, i.e., it matters when we talk about the health of a person that we consider how that person is experiencing themselves and their body (e.g., Carel, 2007, 2018). A philosophical account of health ought to sit comfortably with our experience of *what it is like* to be healthy or unhealthy. Similarly, an account of social disruption that flew in the face of our experience of *what disruption is like* for us should not be regarded favourably. Of course, our experiences can be questioned or reinterpreted– this is true of both health and social disruption– but an adequate account of social disruption ought to be consistent with the best available interpretation of our experience of social disruption.

3. On the flipside, it would capture what it is that makes a phenomenon distinctly socially disruptive, as opposed to an aggregate of individually disruptive phenomena.

This desideratum reflects an important disanalogy between social disruption and medical health. Health is a property of individuals. While we can speak meaningfully of "social health" or of the state of health of a society, what is meant when we speak of such things is the proclivity of a particular social arrangement for promoting individual health and the statistical measure of the general level of individual health across a society, respectively. Health is not an inherently social property. In contrast, social disruption should be understood as an *inherently* social phenomenon. If an account of social disruption fails to appeal to any inherently social processes, properties, or states of affairs, arguably it is not an account of social disruption at all, but rather a change of subject.[4]

4. It would elucidate what it means to overcome an episode of social disruption in a progressive way.

History teaches us that we can recognise after an episode of social disruption has taken place that what used to seem natural and normal– or "healthy"– to us was in fact morally objectionable (Pleasants, 2008). A case in point, which at once exemplifies the disruption and progressive revision of moral and medical norms, is that of the (albeit incomplete) transformation of moral and medical attitudes towards homosexuality that took place in the late 20th century (Baker, 2019, pp. 45–49). Such recognition is possible precisely because social disruption shatters complacency about what counts as the "natural," "normal" way to proceed within a given domain of inquiry. Just as becoming unwell under certain circumstances can prompt one to change one's

---

[4] In contrast, Hopster is explicit that his account of socially disruptive technologies includes disruption at the level of individuals as well: "technosocial disruption need not necessarily pertain to distinctly social practices and relations; it may also pertain, for instance, to the cognition or sensory experience of individuals" (2021, p. 1, n. 1). As my desideratum suggests, I find it strange to include disruptive phenomena with no social component under the umbrella of technosocial disruption.

lifestyle for the better– perhaps even initiating reassessment of what it means to be healthy– so too can social disruption lead to desirable social change, if we respond to it correctly. In constructing an account of social disruption, therefore, one must recognise the fact that episodes of social disruption can be moments of profound ethical vulnerability, risk, and opportunity. It matters ethically that we respond to episodes of social disruption *correctly*. The LGBT/Gay Rights revolution was not mere change– it was *moral progress*. To be clear, a philosophical account of social disruption does not need to determine whether a *particular* response to disruption is progressive or not– this is a job for first-order ethics– but it does at least need to accommodate the *possibility* of progress resolution.[5] If an account of social disruption cannot make sense of this possibility, then we ought not to regard it favourably.

5. Lastly, it would allow us to understand how a socially disruptive phenomenon causally brings about its socially disruptive effects.

Effective medical intervention relies on understanding the causal processes underlying a disease and interrupting or exploiting those processes to prevent, cure, or treat it. If we accept that calling something a social disruption involves a commitment to resolving or overcoming it, we should aspire for our account of social disruption to empower us to do so systematically and sure-footedly. Ideally, it would provide an understanding of the causal mechanisms underpinning a social disruption that would empower us to "get ahead" of social disruption: to explain, anticipate, avoid, intervene in, or exploit future episodes of social disruption in desirable ways. Ultimately, this is the desideratum that my account of social disruption has difficulty satisfying; it characterises social disruption as something with no underlying essence, something of which there is no "getting ahead."

In the next section, I critically survey some of the existing literature on social disruption through the lens of these desiderata. This serves as a springboard for the account of social disruption I will propose in Sects. 4 and 5.

3 Social Disruption and Uncertainty

3.1 Undermining Established Norms and Codes

Robert Baker's influential account of techno-moral disruption is presented as an extension of Clayton Christensen's business-oriented account of "disruptive innovation" (Bower & Christensen, 1995; Christensen, 1997), according to which changes in commercial practices are induced by a technological innovation.[6] Baker writes that

---

[5] Furthermore, this desideratum is metaethically neutral: it is equally well satisfied by realist theories (progressive resolution is greater conformity to objective values) or anti-realist theories (progressive resolution is a collectively assigned social status, for example).

[6] Christensen did not invent the notion of disruptive innovation *ex nihilo*. His explication of the concept of disruptive innovation was influenced by the views and ideas of significant thinkers before him. In particular, there is a lineage connecting the concept of disruptive innovation with that of "creative destruction", most associated with Schumpeter (2013)– the process by which new innovations obsolesce older ones,

a technology is *morally* disruptive to the extent that it forces change in "established moral norms or ethical codes" (2013, p. 59), the supposed counterparts to commercial practices in the moral domain.

As "perhaps the clearest example" of a morally disruptive technology, Baker gives the mechanical ventilator. This is the example I will return to throughout this paper. While there is broad agreement that the mechanical ventilator is indeed an excellent example of a morally disruptive technology, it has been noted that the ventilator is not demonstrative of a perfect analogy between commercially and morally disruptive technologies. Compare the effects of the mechanical ventilator with those of the refrigerator, an example offered by Baker of a commercially disruptive technology. Historically, both the mechanical ventilator and the refrigerator were invented to serve an overarching purpose: the ventilator to enable physicians to extend the lives of those in their care; the refrigerator to keep one's perishables fresh for longer. Now, once you have a fridge, there is no ambiguity about whether the overarching aim of keeping your perishables fresh still requires, for example, daily ice deliveries: it unambiguously does not. This is the sense in which the refrigerator disrupted established commercial practices such as ice deliveries: it clearly obsolesced them. However, in the case of the ventilator, technological innovation created a previously unencountered set of circumstances in which there was a great deal of ambiguity about what was demanded of physicians to satisfy their overarching duty of care. It is the mechanical ventilator's capacity to create such an ambiguous set of circumstances in which its special sort of disruptive power, if we can call it that, resides.

Such considerations are what motivate Nickel, Kudina, and van de Poel's claim that "moral routines can change without being disrupted" (2022, p. 263). Switching examples to the contraceptive pill, they imagine a hypothetical world

> where in one year, birth control pills are not available and people are urged to be chaste and faithful, and in the next year, the pill becomes available, and people celebrate reproductive autonomy. In this imaginary world, the change happens in such a way that people simply adjust to a new way of doing things (ibid.)

In this imaginary world, the effect of the contraceptive pill on the moral norms governing sexual practices is precisely analogous to that of the refrigerator on the commercial practice of delivering ice. They imagine that once the contraceptive pill became available, it was immediately clear to everyone that restrictive, conservative sexual norms were largely redundant for the purpose of preventing unwanted pregnancies, in the same way that it was immediately clear that daily deliveries of ice were redundant for the purpose of keeping one's perishables fresh.

---

resulting in a treadmill effect that sustains economic growth– which in turn was inspired by Marx's ideas about the self-destructiveness of capitalism (e.g., Marx & Engels, 2011). The genealogy of disruptive innovation can also be traced back to such diverse ideas as Hegel's notion that the process of "sublation" of terms or concepts drives the advance of the dialectic and Darwin's notion that novel mutations drive the process of natural selection by forcing existing species to extinction. Thanks to an anonymous reviewer for prompting me to draw attention to these interesting earlier ideas and their connection to disruptive innovation.

Perhaps the authors could have picked a more plausible example: one can perhaps more easily imagine a world in which the advent of clean nuclear fusion entirely obsolesces the reliance on and use of fossil fuels in a wholly uncontroversial way, or one in which the widespread availability of lab-grown meat obsolesces the practice of farming animals for food without any moral ambiguity. It would not be correct (or at least not very illuminating) to describe these imaginary changes in moral routines, happening in a way comparable to that described by the authors above, as socially disruptive. Momentous and consequential, certainly, but not disruptive. Therefore, these authors argue, moral disruption is "not fully explained by a change of moral routines" (2022, p. 264). Rather, *disruptive* changes of moral routine are character-ised or accompanied by an interpretative uncertainty that undisruptive changes of moral routine lack. This is how I will be understanding disruption going forward. The important question is how we ought to understand this interpretative uncertainty.

3.2 Epistemic Uncertainty

One natural interpretation of the uncertainty that characterises or accompanies moral disruption is epistemic, i.e., uncertainty as a straightforward lack of relevant knowl-edge. Nickel suggests that "in situations of moral disruption, people… are caused not to know their own moral obligations, or those of others (at least not determinately)" (2020, p. 259). Disruptive episodes raise (causal) barriers to knowledge that can be either descriptive or normative or both: we might not know how to conceptualise a disruptive situation, or we might not know how to respond to a disruptive situation; because moral concepts can be thick, descriptive and normative ignorance are often entangled.

Returning to Baker's example of the ventilator, the possibility of artificially main-taining lung and heart function despite a permanent lack of brain function deprived physicians and families of descriptive knowledge regarding the status of a patient in such a state as alive or dead. Furthermore, once a patient was hooked up to a mechanical ventilator, it is possible to keep oxygenated blood pumping through their body, thus preventing ischemic injury to the body's tissue and organs. This means that the patient's organs in effect remain viable for transplantation indefinitely. In this context, physicians lack normative knowledge regarding the permissibility of deliberately ceasing the patient's lung and heart function or of using their still viable organs to perform their duty to save the lives of other patients in their care. More-over, it was normatively ambiguous whether it was appropriate for those close to the patient to begin rituals of mourning. The descriptive uncertainty surrounding the status of the patient as dead or alive is entangled with the normative uncertainty sur-rounding the transplantation of the patient's organs and rituals of mourning because the concept of death is thick.

Something Nickel makes clear in several places is that in the relevant sense uncer-tainty "is not just a psychological state"– in contrast, say, to the psychological state of being unsure or lacking confidence– "but an epistemic state in which a person has insufficient reason to close a given question" and "it is this epistemic insufficiency of reasons that partly constitutes uncertainty" (2020, p. 260). Nevertheless, this epis-temic state normally triggers a certain subjective experience, and such subjective

experiences are what Nickel and his co-authors call "the phenomenological core" of disruption (Nickel et al., 2022, p. 277). Therefore, while the epistemic account describes disruption in the language of a lack of epistemic reasons as opposed to a subjective experience of uncertainty, the account does suggest a way of satisfying the second desideratum I proposed in Sect. 1, that of providing an account of the individual phenomenology that accompanies disruption.

Because this lack of epistemic reasons is typically experienced as unpleasant (Nickel, 2020, pp. 261–262), disruption, according to the epistemic account, triggers a process of inquiry among those affected by it in hopes of relieving that unpleasantness. The way Nickel and his co-authors describe the kind of inquiry initiated by disruption suggests a way of distinguishing a healthily functioning moral practice from a moral practice that is undergoing disruption: they describe the inquiry triggered by epistemic uncertainty as "conflicting and sometimes antagonistic" (Nickel et al., 2022, p. 262); as "personal, hazardous, and resource-intensive", in the sense that individuals affected by disruption "have a personal stake" in resolving it and must effortfully exercise their moral imagination to do so (2022, p. 277). Furthermore, there is always the risk that a community's efforts may result in failure, adding a potent sense of precarity and distress to the characteristic experience of disruption. This suggests a way of satisfying the first desideratum I proposed in Sect. 2, that of differentiating between a domain of inquiry that is functioning healthily and one that is experiencing disruption.

However, the epistemic account of disruption just articulated invites two objections. The first is that the kind of inquiry supposedly triggered by and characteristic of moral disruption, far from being an *interruption of* or *obstacle to* healthy moral practice, is *characteristic of* moral practice functioning in perfect health, i.e., in a way that is conducive to progress. Therefore, the epistemic account misidentifies social disruption with moral inquiry functioning in good health and consequently fails to satisfy the first desideratum proposed in Sect. 2. The second objection is that the characteristic features of disruption, on the epistemic account, seem to lack an irreducibly social component that is necessary for a supposed instance of social disruption to be genuinely social in nature. Therefore, the epistemic account also fails to satisfy the third desideratum I proposed in Sect. 2, that of accounting for the inherently social nature of social disruption. I will now elaborate on these objections in more detail; they will prompt me to shift my focus to the phenomenon of "deep" disruption.

3.3 Doubts about Epistemic Uncertainty

With regards to the first objection to the epistemic account, my central claim is that the "natural," "normal" flow of moral inquiry – that is, moral inquiry functioning healthily, in a way that is conducive to progress – does (or should) involve often heated exchanges between proponents of a diversity of views, opinions, and perspectives, and that the discomfort associated with such confrontations is part and parcel of participation in healthy moral inquiry. Consequently, the epistemic view misattributes to moral disruption what is in fact characteristic of healthy moral inquiry. To substantiate this objection, I am going to draw on an existing, well-known debate from the philosophy of science, regarding the proper roles of diversity of opinion and

heated, often personal, disagreement in progressive scientific discourse. The analogy with scientific discourse is all the more relevant because several of the authors who have contributed to the formulation of the epistemic view of social disruption (Baker, 2019; Nickel et al., 2022, pp. 265–266) have been inspired by (a certain reading of) a controversial position within this debate in the philosophy of science– namely, Thomas Kuhn's (2012) picture of scientific inquiry as consisting of stretches of consensus punctuated by moments of controversy, crisis and subsequent revolution.

In opposition to this reading of Kuhn's philosophy of science, Imre Lakatos regarded competition between rival scientific perspectives as one of the core drivers of scientific progress and a ubiquitous feature of historical and contemporary scientific practice. He writes:

> *The history of science has been and should be a history of competing* [*scientific perspectives*] *… but it has not been and must not become a succession of periods of normal science: the sooner competition starts*, *the better for progress.* (1970, p. 155)[7]

This Lakatosian view is elaborated upon by Kitcher, for whom a diversity of scientific perspectives "serves the community as a hedge against becoming stuck" (1995, p. 72); "a community that is prepared to hedge its bets when the situation is unclear is likely to do better than a community that moves quickly to a state of uniform opinion" (1995, p. 344).

For precisely these pragmatic reasons, Kitcher also views such diversity of opinion as a healthy feature of moral inquiry as well. He envisions a "Deweyan society" that "encourages some of its members to think about alternatives" (2021, p. 97) to even the most established moral norms in hopes of identifying overlooked refinements to our moral practices thus creating opportunities for progress (2021, p. 67). I have written elsewhere about the extent to which it makes sense to regard ourselves as obligated to consider such unobvious alternatives (Lane, 2023), but for present purposes, it seems clear that a diversity of perspectives and the existence of conflict are signs of a domain of inquiry in *good* health, i.e., functioning in a way that is conducive to progress.

Furthermore, a scientific community in which proponents of rival perspectives are prepared to (often vociferously) defend their views in the face of opposition is more likely to experience scientific progress than one in which they are not, since drawn-out exchanges between rivals, in which competing theories are articulated in ever more precise terms and put to the test in ever more demanding contexts, has the iterative effect of refining the ultimately successful view. As Popper put it:

> The dogmatic attitude of sticking to a theory as long as possible before giving it up is of considerable significance. Without it we could never find out what

---

[7] Lakatos here is indebted to Popper, who wrote that the "struggle of ideologies… seems to be characteristic of anything that might be called development in human thought" (2007, p. 420). Lakatos's contribution to the Popperian view was to synthesise Popper's emphasis on struggle with Kuhn's notion of paradigms, replacing the picture of competition between individual theories and hypotheses with that of competing "programmes" (Lakatos, 1978).

was in a theory– we should give the theory up before we had a real opportunity of finding out its strength; and in consequence no theory would ever be able to play its role of bringing order into the world, of preparing for future events, of drawing attention to events we should otherwise never observe. (2007, p. 420, n. 1)

The history of science contains plenty of examples of the dogged persistence of rivals ultimately serving to improve the eventually emergent perspective (Kitcher, 1995, Chap. 6). It is easy to recognise that the dogmatic attitude plays the same role in moral discourse as it does in scientific discourse, namely, to give a moral "theory" (a particular conception of a moral concept, a particular interpretation of a moral principle or value, etc.) the best opportunity to "play its role" in our lives: to order our experience, to prepare us to handle future situations, and to draw our attention to significant phenomena that we might otherwise overlook. Therefore, the "conflicting and sometimes antagonistic" discourse (Nickel et al., 2022, p. 262) that Nickel and his co-authors identify with disrupted moral inquiry is in fact an epistemologically useful feature of moral inquiry functioning healthily.

Beyond the benefits of conflict and antagonism, some philosophers regard contention, in which participants in a discourse have a personal, emotional stake, as epistemologically essential to moral inquiry as well. Anderson suggests, also along Deweyan lines, that

knowledge of the right arises from certain distinctive experiences: of being held subject to the claims of others, which are asserted as authoritative, of being called to account by others, held responsible, blamed, shamed, and punished for wrongdoing, being criticized for arrogance, negligence, and other vices, being exposed as unable to justify one's conduct before others whom one has injured or neglected. (Anderson, 2014, pp. 7–8)

Furthermore, elsewhere she suggests

that emotions help us focus on normatively relevant features of urgent problems, and that the lack of emotion of the privileged may reflect their indifference to the plight of others.… *lacking stakes…* may make moral reasoning irresponsible and unaccountable to those to whom the outcomes matter. (Anderson, 2015, p. 26)

In other words, emotional contention in which participants "have a personal stake" (Nickel et al., 2022, p. 277) is epistemically essential to effective moral inquiry, and not a disruption to it, as Nickel and his co-authors seem to suggest. Moral inquiry in which participants lack emotional stakes is epistemically impoverished and unlikely to lead to progress, biasing inquirers with the greatest sway towards a status quo that serves their own interests.

Returning to the analogy with scientific discourse, as Kuhn was at pains to make clear, among their motivations for doing science, scientists have a professional stake in its (and, within it, their own) success; furthermore, the resolution of scientific dis-

agreement is resource intensive, often demanding enormous feats of scientific imagination from scientists. Moreover, scientists' personal investment in scientific issues is *good for* science: emotionally invested scientists are unlikely to give up on theories in which they have invested time and effort to develop, resulting in the kind of attention to detail and constructive competition between rival theories described above.

In addition, there is always the risk that the efforts of the scientific community may result in failure. As Popper recognised, "there are many instances of futile struggle in the history of thought, struggles which have ended in nothing" (2007, p. 424). Personal investment, resource intensiveness, and some personal and professional precarity are the rule, rather than the exception, for academic scientists. While *moral* agents' motivations for participating in moral discourse are very different from scientists' motivations for doing science, if Anderson is right about how moral discourse does and should work, something similar goes for them too: personal stakes, demandingness, and some personal precarity are indicative of properly functioning moral inquiry, not a disruption to it.[8] Once again, the need to engage in "hazardous and resource-intensive" (Nickel et al., 2022, p. 277) exploration of alternatives is not a disruption to healthy moral inquiry but a feature of it.

My suspicion is that Nickel and his co-authors' misidentification of healthily functioning moral practice with moral disruption can be attributed to an attempt to be even more thoroughly "Kuhnian" in their characterisation of disruption than Baker was (Nickel et al., 2022, pp. 265–266). This might explain why they make the common mistake among those inspired by Kuhn of neatly dividing the activity of inquiry into stable periods of largely habitual puzzle-solving punctuated by disruptive crises of confidence and subsequent revolutions– precisely the picture of science disavowed by Lakatos and subsequent philosophers of science (and which I regard as a stemming from a misreading of Kuhn). But moral inquiry is not and should never be an activity of emotionally detached puzzle-solving. If a degree of public disagreement and private discomfort is characteristic of moral inquiry functioning healthily, in a way conducive to progress, then we need an account of what distinguishes undisruptive uncertainty from disruptive uncertainty,[9] and for this we must go beyond the epistemic account. There is more to social disruption than being in a state of needing additional hints or clues to solve a puzzle.

My second objection to the epistemic account is that it fails to capture a crucial qualitative feature of social disruption: its irreducibly social element– the third desideratum proposed in Sect. 2. Nickel and his co-authors are somewhat equivocal about the relationship between the individual and social components of disruption on the epistemic account. On the one hand, Nickel writes that the kind of moral uncertainty

---

[8] As I state in the main text later, I do not want to be unreasonably strict about what gets called disruption and what doesn't; conflict, of the kind emphasised by Nickel, Kudina, and van de Poel in their discussion of inquiry triggered by disruption, is clearly disruptive in a familiar, intuitive sense. Rather, the kind of disruption that I am interested in is disruption of *the healthy functioning of a domain of inquiry*, within which conflict can and does have a proper place. The kind of conflict described by Nickel, Kudina and van de Poel might then be regarded as a form of "healthy disruption" (if such a thing is not oxymoronic). I do not mind which way the issue is phrased.

[9] Or, if you prefer, we need an account of what distinguishes "healthy disruption" from unhealthy disruption.

characteristic of moral disruption "is itself a social phenomenon, transcending the level of individual psychological explanation" and that consequently "we are in a condition of moral disruption together" (2020, p. 262). This sounds right to me and will be a feature of the rule-following account I begin to develop later. On the other hand, Nickel and his co-authors write as if the relevant kind of uncertainty is an individual phenomenon that snowballs into aggregate uncertainty at the social level: uncertainty "will typically start individually, but it can also become collective if an entire group, e.g., of practitioners, becomes uncertain about what values to apply or how to apply them" (Nickel et al., 2022, p. 261); moral disruption

> starts as an individual phenomenon and is about being uncertain, confused or perplexed about what values and norms to apply or how to apply them. Groups can be said to be morally uncertain when their members are individually or mutually uncertain. (2022, p. 274)

In the context of the quoted passages, our being "in a condition of moral disruption together" reads more like a quantitative criterion– that enough individuals in a society experience the disruption– and not a qualitative one, that the disruption be *distinctly* social in character.

Similarly, these authors write that while moral inquiry triggered by moral uncertainty "may manifest itself individually, it is typically collective in nature as it is aimed at (re)establishing community consensus on how to deal with a certain moral issue" (Nickel et al., 2022, p. 274). They continue:

> While inquiry often seems to start individually, it is a collective or communal activity whose success in overcoming initial uncertainty requires the relevant social group, or groups, to accept it as a solution to the initial moral unease. (2022, p. 278)

It is not clear to me what the authors mean when they write that the resolution of disruption requires *the group* to accept a proposed solution. Does the group accept as a collective or as a collection of individuals? The resolution of collective moral uncertainty through inquiry these authors have in mind reads to me, in the context of my reading of their view of the uncertainty brought on by social disruption, like the resolution of *individual* uncertainty writ large. This interpretation, I think, omits one of the most important features of the resolution of social disruption as *a distinctly and irreducibly social achievement*.

To be clear, I'm not interested in being overly strict about what can and can't be labelled "disruption"; obviously, an insufficiency of epistemic reasons (and the bickering that ensues) *is* disruptive in an important sense, because such insufficiencies can get in the way of doing things that badly need to get done. Rather, I would prefer to be interpreted as arguing that the epistemic account is not a *complete* account of disruption, because it fails to capture an especially philosophically interesting *kind* of disruption, the kind which I think deserves the label "social disruption."

4 Social Disruption and the Possibility of Inquiry

4.1 "Deep" Disruption

I have argued that the discursive and phenomenological features that Nickel and his coauthors associate with moral disruption are in fact characteristic of moral inquiry functioning healthily, i.e., in a way that is conducive to progress– and, consequently, that the account fails to satisfy the first and second desiderata I proposed in Sect. 2. As these authors point out, the kind of inquiry they have in mind "depends on the ongoing existence of a conversation or community discourse in which consensus remains a possible outcome" (2022, p. 275). A pertinent question for understanding social disruption is therefore: what supports this possibility?

I am going to argue that the possibility of the kind of disagreement characteristic of healthy moral inquiry is, somewhat counterintuitively, made possible by the existence of a very particular form of agreement among disagreeing parties, what we might call "deep agreement."[10] Deep agreement is pre-epistemic. Deep agreement supports both agreement and disagreement at the epistemic level. Deep agreement makes it possible to exchange and assess the claims and arguments over which we agree or disagree. Social disruptions, I am going to argue, are occasions for the breakdown of such deep agreement. The phenomenological core of social disruption is what McDowell once described as the sense of "vertigo" (1981, p. 149) experienced when the ground provided by this agreement is pulled from under our feet (or when we become aware of the instability of that ground). Furthermore, such agreement is *irreducibly* social in nature, and its restoration is an irreducibly social achievement. Thus, from the outset, a "deep" or pre-epistemic account of social disruption suggests promising ways of satisfying the first, second, and third desiderata I proposed in Sect. 2 – something the epistemic account was unable to do. I will elaborate on these in the remainder of this section. However, as I will demonstrate later, the most philosophically interesting consequences of my account have to do with its ability to satisfy (or not) the fourth and fifth desiderata: to elucidate what it means to overcome an episode of social disruption *progressively*, and to explain how a socially disruptive phenomenon, such as the innovation of a SDT, brings about its socially disruptive effects.

The notion of deep disruption is not new. Hopster characterises deeply disruptive technologies as those that "make us lose our normative, theoretical and conceptual bearings" and initiate "a reassessment of basic building blocks of ethical theorizing" (2021, p. 4). An understanding of deep disruption as an interruption or perturbation of the social conditions that support the possibility of reaching determinate answers to moral questions has recently been articulated by Hermann (2022, 2025).

Among the basic building blocks of ethical theorising vulnerable to technological disruption are our conceptual practices. According to Löhr, conceptual disruptions challenge

---

[10] We might also call it agreement "in form of life" (Wittgenstein, 2009, §241), "in judgement" (2009, §242, p. 238), or perhaps most perspicuously, agreement "in action" (Wittgenstein, 1996, VI §39).

the ways in which an individual or group has intuitively classified individuals, properties, actions, situations, or events, leading to a classificatory conflict or uncertainty, i.e., a conflict or uncertainty about the application conditions of a word or concept. (2023a, p. 3)

Marchiori and Scharp (2024) characterise conceptual disruptions as interruptions to the normal functioning of an individual concept, cluster of concepts, or whole conceptual scheme. Indeed, Löhr (2023b, pp. 4–5) had earlier suggested what it really means for an individual concept to change is for the relations between it and the other concepts surrounding it in its conceptual network to change. It is these conceptual networks, considered as wholes, that constitute our conceptual resources, not the sum total of our concepts considered in isolation. Conceptually disruptive technologies, according to Löhr, can "generate new artifacts, states, or events that we currently are not in the position to clearly classify given our current conceptual resources" (2023a, p. 6).[11] Clearly, the mechanical ventilator can be regarded as a conceptually disruptive technology in this sense: physicians and families were not "in the position", whatever this means, to clearly classify the state of a patient hooked up to a ventilator as determinately alive or dead, nor to clearly classify the act of "pulling the plug" on such a patient as an act of, say, murder, justified killing, allowing to die, or simply switching off bodily support for an already (brain-)dead patient. This resulted in a classificatory conflict or uncertainty that is at once descriptive and normative. But this was not an epistemic problem, solvable by accumulating further epistemic reasons, something an individual could in principle do by themselves. Rather, according to Löhr, the resolution of classificatory ambiguity demands "a *collective decision* as to how to classify an event or object" (2023a, p. 6),[12] in this case, the articulation and collective recognition of the concept of brain-death and its significance as death *of the person* as opposed to (merely) death *of the body*.

A helpful way of bridging the notion of conceptual disruption with that of disruption of effective inquiry more broadly is provided by Queloz's broad characterisation of concepts as "thinking techniques: as the norm-governed patterns according to which we move from perception to thought, from thought to thought, and from thought to action" (2021, p. 23). The notion of thinking techniques broadens the concept of social disruption to cover ambiguities beyond the classificatory. We regard

---

[11] Marchiori and Scharp wonder whether some cases of conceptual disruption might not be classificatory in nature (2024, p. 18). As a counterexample, they appeal to the "disruptive" effect of quantum mechanics on logic, prompting the development of non-classical (quantum) logics (e.g. Putnam, 1979). I am not sure whether logical disruptions, if such things are even possible, should count as conceptual disruptions; it might be more apt to describe them (hypothetically) as belonging to the family of disruptions to the thinking techniques constitutive of moral competence (see immediately below). However, I am not convinced by the counterexample, since "quantum logics" never really gained enough steam to qualify as a real disruption– and for good reason. For an entertaining and informative account of the history of quantum logics, see (Maudlin, 2005).

[12] The language of "collective decision" risks misrepresenting the process of settling on a new conceptual practice as that of rounding up reasons and making a choice on the basis of them. In some contexts, the simpler language of a community coming to "collectively recognise" a particular classification may be more apt. I have discussed the variable extent to which the articulability of reasons matters when it comes to settling conceptual disputes in a progressive way (Lane, 2024).

as a morally competent agent someone who has mastered the thinking techniques of moral inquiry, including but not limited to its norms of classification. Thus, episodes of social disruption can interrupt or perturb *not only* our classificatory practices *but also* the norms of thought that govern the inferences we draw and the actions we take on the basis of our classifications. Along these lines, Hermann describes moral competence "as a complex capacity involving emotional as well as reasoning skills, habits, and dispositions" (2023, p. 21). Possession and reciprocal recognition of moral competence, including moral-conceptual competence, by and between members of a community of moral inquiry is "what makes practices such as morally justifying, morally praising and blaming, asking for moral reasons, etc., possible" (p. 32). Deep disruption, then, can be understood as challenging in the broadest sense "what moral competence involves" (p. 19) in a particular context, thus calling into question what successful moral inquiry even amounts to. Thinking about disruption in terms of the social conditions that make moral inquiry possible therefore gives us a way of differentiating between a disrupted moral practice and a moral practice in good health–something the epistemic account was unable to provide.

It is also capable of capturing the "phenomenological core" of disruption just as well as the epistemic account, if not better. As Löhr points out about conceptual disruption, "[l]eaving a situation or artifact unconceptualized can generate not only communicative disruption but insecurity, distress, and uncertainty in the community and its members" (2023a, p. 3). Hermann, discussing the effect of care robots on interpretations of moral competence in the context of care homes for the elderly, writes that caregivers and -receivers will feel untethered from the tacit practical foundations on which they had previously based their care practices, coming to doubt

> what it means to be a good caregiver, what good care is, how moral responsibilities are distributed among themselves and the robots, what they owe to the elderly and what the moral status of the robots is. (2022, p. 27)

To reiterate, what is required to alleviate the insecurity, distress, and uncertainty experienced by human caregivers and the elderly in this context is *not* the further accumulation of epistemic reasons, something that an individual caregiver or -receiver can in principle do by themselves, but a collective decision about *what competent care involves in this context*, a decision that will subsequently enable the caregivers and -receivers to answer the question of who owes who what in this context. The source of disruption is not a hindrance to the exercise of moral competence but rather a lack of consensus among the relevant community about what moral competence involves in a new context; because such consensus is what supports our individual practices of inquiry, its absence results in the subjective experience of insecurity and disorientation. The unpleasantness of such subjective experiences provides the private impetus to reestablish public consensus.

In what immediately follows, I am going to develop this view of social disruption further, articulating an underappreciated connection between disruption, rule-following, and the possibility of progress through inquiry. On the basis of the extended discussion of rule-following in Sect. 5, I conclude that an adequate understanding of this connection makes it extremely difficult to understand the causal connection

between socially disruptive phenomena and their socially disruptive effects– the fifth desideratum I proposed in Sect. 2. Accordingly, I suggest that we ought to adjust our expectations of what an account of social disruption and SDTs can deliver.

4.2 Inquiry, Progress, and Rule-Following

To the extent that we make moral progress through moral inquiry– something that some authors have recently expressed agnosticism about (e.g. Rehren & Blunden, 2024)– we do so by taking our moral concepts, principles, and values, and appropriately bringing them to bear on moral problems that we have either previously neglected, overlooked or not encountered before. Our moral concepts, principles, and values cannot shape our response to a moral problem in just any arbitrary way (we must do more than pay them lip service); there is a right and wrong way to bring them to bear on our problems. In other words, within the activity of moral inquiry, our moral concepts, principles and values are subject to standards of correct use.

The relevant sense of correct use here is not a one-off correct use (e.g., a correct guess) but a robust, temporally extended capacity to use correctly– what we might want to call a correct *grasp* of the concept, principle, or value. The kind of moral progress we are interested in is what Schinkel and de Ruyter call "moral progress in the strong sense" (2017, p. 122): a non-superficial, durable change for the better. It couldn't be said that I was able to ride a bike until I could *go on* cycling; likewise, it couldn't be said that I had become competent with a moral concept, principle, or value until I could reliably *go on* using it in a variety of different and new situations.

In *On Certainty*, Wittgenstein writes: "Certain events would put me into a position in which I could not go on with the old language-game any further. In which I was torn away from the sureness of the game" (2008, § 617). The impact of technology on our ability to "go on any further" in our moral inquiries– its capacity to "tear us away from the sureness of the game"– has received some limited attention in the literature (e.g. Hermann, 2022, p. 22; Löhr, 2023a, pp. 3, 6). It has, for example, been recognised that "human thinking and acting can be technologically mediated" and therefore that technological innovation can "create new relationships and obligations, change practices fundamentally, and create new epistemic and moral landscapes" (Hermann, 2022, p. 32) that affect our ability to go on in the same way with our moral inquiries.

However, except for Hermann, who rightly emphasises the role of "moral certainties" in scaffolding effective inquiry, not much attention has been paid to what exactly it means for moral inquiry to "go on in the same way". In *Philosophical Investigations*, Wittgenstein also wrote that "[t]he use of the word 'rule' and the use of the word 'same' are interwoven" (Wittgenstein, 2009, § 224). Accordingly, I propose that we understand this ability to "go on in the same way" in terms of our ability *to follow the rules* constitutive of the thinking techniques that are the engine of moral inquiry; the question of what makes it the case that we are going on in the same way therefore

becomes that of what makes it the case that we have followed or are continuing to follow the relevant rules.[13]

5 Two Pictures of Rule-Following; Two Pictures of Social Disruption

In this section, I contrast two plausible pictures of rule-following and draw out their implications for an account of social disruption. The first picture to be discussed– the "rules-as-rails" picture– is a prima facie plausible view of rule-following that coheres with casual (philosophical) talk of "going on in the same way" as grounded in an individual's intentions regarding and understanding of their own rule-guided actions. To be clear, I do not attribute this picture to any existing accounts of social disruption in the literature. I consider the rules-as-rails picture as a prima facie plausible *interpretation of* the account of social disruption developed by Hopster and Löhr, but do not consider it to be *part of* their account. In the context of my argument, the point of discussing this "common sense" interpretation first is to demonstrate that prima facie plausible ideas lead to certain surprising difficulties if we want to understand the connection between rule-following and social disruption in a way that optimally satisfies the desiderata proposed in Sect. 2. In particular, the rules-as-rails picture cannot make sense of the idea of responding *correctly* to an episode of social disruption and therefore cannot make sense of the idea of overcoming an episode of social disruption in a progressive way– thereby failing to satisfy the fourth desideratum proposed in Sect. 2. This motivates the development of an alternative picture– the "language-game" picture– that, while perhaps less intuitive, is both more philosophically sound in its own right and avoids the difficulties one runs into if one accepts the "rules-as-rails" picture. However, I will demonstrate that even this superior account of rule-following in the context of social disruption cannot satisfy all of my proposed desiderata either. In particular, it fails to satisfy the fifth desideratum: it cannot make sense of how socially disruptive phenomena, such as SDTs, bring about their own disruptive effects, thereby disempowering us from anticipating and mitigating social disruption in desirable ways. I ultimately conclude that we should reassess what it is reasonable to expect a philosophically sound account of social disruption to deliver.

Throughout this section I draw heavily on Wittgenstein's rule-following considerations and the literature that surrounds them. I am not offering a Wittgenstein exegesis; I do not claim that this is what Wittgenstein himself did or would say about the phenomenon of social disruption. Besides, what Wittgenstein did or would think and what one has good reason to think are two separate questions (often with different answers). Rather, I offer an account of social disruption that is inspired by a particular reading of Wittgenstein, and my discussion below wears that influence on its sleeve by frequently referring to Wittgenstein's writings, especially in *Philosophical Investigations*, as well as to the literature surrounding it. However, it is important that my account can stand on its own two feet. I therefore invite and encourage the reader to engage critically with my account, not on the basis of whether it is properly Wittgen-

---

[13] I intend this approach to supplement and complement Hermann's account in terms of certainty, though I don't have the space to fully flesh out the connections between them in this paper.

steinian (although this is an interesting interpretative question), but on the basis of the soundness of my argument for it.

5.1 The Rules-as-Rails Picture

How should we understand the claim that an individual has "gone on in the same way," i.e., has followed or is continuing to follow a particular rule?

One intuitive answer centres on the idea that the correctness of an individual's rule-following activity is "in a *strange* way" (Wittgenstein, 2009, §195) fixed or set by their "grasp" of a rule. When we grasp a rule, according to this view, we commit ourselves to the standards of correct use that govern its future application. As McDowell puts it, borrowing a Wittgensteinian metaphor (2009, §218), an individual's grasp of a rule "is pictured as something like engaging mental wheels with objectively existing rails" (1981, p. 146). Once the wheels of a train engage with the tracks, the train is kept going in the right direction (in the absence of external countervailing force, e.g., a fallen tree or a landslide). Likewise, once an individual has grasped a rule, their grasp of it will correctly determine their future rule-following activity (in the absence of external countervailing forces). Such states of grasping can be invoked to explain why groups of individuals go on in the same or similar ways in their rule-following practices: just as different trains invariably reach the same destination when they engage with the same rails, so too do different individuals "go on in the same way" when they grasp the same rule. On this picture, a community's shared grasp of the same rules explains the deep agreement that makes inquiry possible.

According to rules-as-rails, social disruption occurs when the proliferation of new circumstances outruns a community's mutual grasp of the relevant rules, whether conceptual or deliberative. This is one way to understand Löhr's claim that conceptually disruptive technologies create circumstances "that we currently are not in the position to clearly classify given our current conceptual resources" (2023a, p. 6)– the resources in question being our grasp of the relevant rules for applying our concepts. The rules-as-rails picture would extend this claim to include our deliberative resources more generally. For example, on the rules-as-rails picture, the circumstances created by the innovation of the mechanical ventilator can be regarded as beyond physicians' imperfect grasp of the relevant rules for classifying patients as dead or alive and of the thinking techniques for bringing such classifications to bear on their practical choices regarding organ transplantation. On the other hand, in the imaginary world considered by Nickel, Kudina, and van de Poel, the relevant community's shared grasp of the relevant rules for competently judging permissible sexual activity might be regarded as sufficiently rich to accommodate the circumstances created by the introduction of the contraceptive pill. (Ditto for my imaginary examples of judging the right ethical response the circumstances created by the advent of nuclear fusion or to lab-grown meat.) Disruption can thereby plausibly be differentiated from healthily functioning inquiry in terms of grasping.

The rules-as-rails picture also captures the phenomenology of rule-following well, as Julian Dodd observes:

[C]onceiving of one's grasp of a rule in terms of being engaged on set of rails that, so to speak, binds one to a pattern of future use does justice to the phenomenology of 'blind' rule-following (1953: §219). It nicely captures the feeling of unquestioning security one feels when one follows a rule with complete assurance, without feeling any need for an inner, mental justification (2023, p. 95).

If we understand the phenomenology of social disruption as the *absence* of such unquestioning security and complete assurance, and the presence of a felt need for an inner, mental justification for how we proceed when confronted with a new situation, then the rules as rails picture suggests a way of capturing the phenomenology of being a participant in a practice that is socially disrupted.

The rules-as-rails picture suggests a *prima facie* plausible way of satisfying the fourth desideratum: that of making sense of the notion of overcoming social disruption progressively. Social disruptions, according to rules-as-rails, are resolved when we mutually improve our grasp of the relevant rules– e.g., coming to better appreciate the "semantic depth" of our existing moral concepts (Moody-Adams, 1999, p. 169)– and *re*engage mental wheels with the objectively existing rails, extending out to infinity, that had guided our inquiry in previous contexts, i.e., when our rule-following activity *clicks* back into place. This can happen simply by coming to appreciate how our current concepts ought to have been brought to bear on a new situation all along, or by becoming more competent in the relevant thinking techniques for intelligently refining our current concepts or innovating new ones in light of the new situation. As Miranda Fricker recognises, "[c]onceptual resources are resources for generating infinitely many *new* meanings, whether as applications of old concepts or coins of new concepts" (2007, p. 104). Both ways involve mutually improving our grasp of rules, whether they govern our practices of conceptual categorisation or the thinking techniques characteristic of competent moral inquiry, including competent conceptual engineering. For example, in the case of the mechanical ventilator, resolution might be regarded as the result of physicians' improving (with philosophers' help) their grasp of the relevant rules to encompass these circumstances, specifically, coming to recognise and make practical use of the concept of brain death.

Lastly, the rules-as-rails pictures seems to satisfy the fifth desideratum as well: that an account of social disruption be able to explain how a socially disruptive phenomenon, such as an SDT, brings about its socially disruptive effects, thereby empowering us to "get ahead" of social disruption, in the sense of equipping up to anticipate and mitigate it in desirable ways. According to rules-as-rails, social disruption is causally explainable psychologically, in terms of the limits of our grasp of the relevant rules. One way of getting ahead of social disruption is by proactively *improving* our grasp of the relevant rules through morally motivated conceptual engineering– what Veluwenkamp et al. understand as the activity of "trying to find out what the correct conception of a concept is in a specific context" (2022, p. 71).[14] On the rules-as-rails

---

[14] Veluwenkamp et al. understand the question of which is the "correct" conception of a concept as that of "which conception is best in line with an ameliorative inquiry, i.e. which conception would grasp best our legitimate and critical purpose of evaluating the moral, political and social dimensions of current specific tech-mediated contexts" (2022, p. 71). They favour a pragmatic interpretation of "best", fixed by the relevant goods and goals that hold sway in a particular context. It is an interesting question how the pragmatic

picture, there is nothing to stop us doing this work in advance of any actual instances of social disruption, and therefore the fifth desideratum appears to be satisfied.

5.2 Derailing Social Disruption

Unfortunately, the rules-as-rails picture of rule-following cannot make sense of the idea of improving our grasp of the relevant rules so as to bring a new situation under them in a progressive way and therefore fails to satisfy the fourth desideratum proposed in Sect. 1.

First, our access to the relevant standards of correct use, against which we judge whether an interpretative decision in a new context *counts as* going on in the same way, presupposes our possession of the very thing we supposedly lack: a grasp of the rule or rules that covers the situation that prompted us to look for an improvement in the first place. Faced with a new situation, we apply our rules in a particular way and say: "this is how the rule is meant to be followed in this context." But is it? What makes it so? We need a meta-rule for correctly interpreting the rule in the new context– and a meta-meta-rule for correctly interpreting that meta-rule, and so on. And this is exactly what we don't have according to the account of social disruption just articulated– these are precisely the resources we lack. Consequently, our judgement that a particular interpretative decision in a new context counts as an improvement "hangs in the air along with what it interprets, and cannot give it any support" (Wittgenstein, 2009, §198).

Second, our interpretative decisions are properly subject to justification. We are not, say, trying to *predict* how physicians will end up responding to the possibilities created by the mechanical ventilator, but to determine how they *should* respond to them, i.e., to determine what would count as a justified response. Even if we grasped the relevant interpretative (meta-)rules, an individual's own mental states (such as their grasp of a rule) are not in the business of justifying the decisions they reach about how to extend their own grasp of a rule to cover a new context. If the standards of correct use against which *I* judge whether an interpretative decision of mine counts as going on in the same way are somehow fixed or set by *my own* grasp of the relevant rules or meta-rules (whatever this means), then "whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'" (Wittgenstein, 2009, §258). For example, even if physicians' mutual grasp of the relevant rules or meta-rules could sensibly be invoked to *explain* their interpretative choices regarding the concept of death in the contexts created by the mechanical ventilator, say, in terms of the dispositions they have as a result of grasping the relevant rule to the extent that they do, such appeals shed no light on why they *should* have extended the concept in the way they were disposed to do. Normative standards of correct use cannot meaningfully be fixed this way, in just the same way that my right hand can-

---

not meaningfully give my left hand money (2009, §268): the relevant institutions are lacking (2009, §380).

The situation I have described with regards to social disruption closely parallels that considered by Saul Kripke in *Wittgenstein on Rules and Private Language* (1982): episodes of social disruption are occasions for the same sort of anxiety about how to go on following a rule that Kripke (inspired by, though perhaps not following, Wittgenstein (cf. Baker & Hacker, 1984)) considers at length. And while many philosophers are comfortable to dismiss Kripke's divisive plus/quus problem as recondite and immaterial, in the case of responding to SDTs, the issue is highly salient and very real: it is a concrete problem with material implications worth taking seriously. *How should we go on* in a moral landscape radically transformed by a new technology? *What does it mean* to respond correctly to such a situation?

The issue is even more disconcerting when we realise that the inadequacy of the rules-as-rails picture is not limited to what we all recognised as disruptive contexts: as Kripke demonstrated, worries about what counts as going on in the same way are articulable even in extremely mundane, totally undisruptive contexts, such as when computing sums like $57+68$.[15] The situations that we call socially disruptive are simply those in which the sceptical doubts articulated by Kripke are *taken seriously*. In undisrupted contexts, we feel confident about how to go on and *just get on with it*. But the rules-as-rails picture provides no grounds whatsoever for this sense of confidence; by itself, confidence is not sufficient for justification. So the rules-as-rails picture of rule-following can't even make sense of the idea of going on in the same way in *un*disrupted contexts; it can't even make sense of what it is we don't have in contexts of disruption. Rules-as-rails therefore fails to satisfy not only the fourth but also the first desideratum proposed in Sect. 2: that of adequately distinguishing between a disrupted practice and a practice functioning in good health.

5.3 The Language-Game Picture

The way out is to give up on the picture of rules-as-rails, and with it the possibilities of making sense of both deep agreement in terms of mutual grasping and social disruption in terms of the limits thereof. An alternative picture– the "language-game" picture– invites us to look beyond the internal state of an individual and to situate an individual's capacity to "go on in the same way" within the context of the patterns of judgements exhibited by the community to which they belong. According to this picture, standards of correct use come not from an individual's own grasp of the rules of a conceptual or deliberative practice, but from the tacit criteria for judging what counts as following a rule, or who counts as a competent rule-follower, operative

---

[15] In *Wittgenstein on Rules and Private Language*, Kripke stipulates that the computing individual has never summed numbers larger than 57 (the choice of 57 is arbitrary)– as stipulated, his question is how this naïve computer can know that their answer (125), and not some other answer (5), is the right one given their previous calculations, the correctness of which are taken for granted. This question mirrors the one about our ethical response to social disruptive technologies. However, the fact is, everyone reading Kripke *has* done sums larger than 57, and the fact that Kripke can make his readers wonder why 125 and not 5 is the right answer is sufficient to show that the rules-as-rails picture is ill equipped to assuage such worries even in undisrupted contexts. The "sceptical paradox" is fully general in its scope.

within a community. So construed, "going on in the same way" in a new context is *not* an intrinsic fact about what an individual is doing or could be doing, say, correspondence between their rule-following intentions and their rule-following activity. Rather, the fact of the matter about whether we are going on in the same way is "observer relative" in Searle's (1995, p. 12) sense; "going on in the same way" is a social status attributed according to tacitly agreed-upon assertability criteria.

It follows that facts about a given situation's potential for social disruptiveness are *not* intrinsic facts having to do with which situations an individual or group of individuals are in a position to handle given the conceptual and deliberative resources at their disposal, i.e., the limits of their grasp of the relevant rules. Rather, social disruptiveness is a status attributed to a situation *in lieu* of tacit agreement regarding the assertability criteria for what counts as going on in the same way in that situation.[16] There are all sorts of explanations for why a community might fail to agree, but the important point is that an absence of deep agreement is only noticeable as such when it gets it the way of something that the community as a whole has set out to do together. Conceptual disruption is not the result of an objective defect of a given conceptual scheme. It is an *inability* of a community to deploy its concepts *to the relevant level of precision*, where the relevant level of precision is fixed by "the relevant projects, goals, and aims of the people involved" (Marchiori & Scharp, 2024, p. 18; cf. Wittgenstein, 2009, §71). Whether one will see a particular episode of interpretative ambiguity *as a disruption* will be highly dependent on one's context, including one's relevant projects and goals.[17]

Furthermore, according to the language-game picture, there is no level of explanation lower than that of the community's tacit agreement for why our rule-following activity is generally harmonious; our deep agreement cannot be explained in terms of mutual grasping; there is nothing that can do the job the rules-as-rails picture invoked grasping to do. And without grasping, there is nothing to assure us that our deep agreement will persist into the future (cf. Kripke, 1982, p. 97). This is perhaps the most disquieting feature of the language-game picture: there are no "objectively existing rails" guiding our interpretative choices even in undisrupted contexts; this is why it is tempting to say that we follow the rules of our conceptual and deliberative practices *blindly* (Wittgenstein, 2009, §219). Given this blindness, it is quite astonishing that we do not lose our normative, theoretical, and conceptual bearings more often than we do. Without grasping, our only recourse is to the language-game that sustains itself because of the brute fact that we generally agree about how we ought to go on in new contexts. But nothing guarantees that this agreement won't break down tomorrow– and with it the possibilities of effective moral inquiry and moral progress. When this happens, we call it social disruption.

Of course, given the fact of our track record of agreement on how to go on regarding a great variety of things, I may reasonably believe that we are in such a position

---

[16] Deep disruption is not a *lack* of rules, otherwise we would not experience it as a disruption, but simply as a free choice. We experience it *as* a disruption precisely *because* we feel like there is "a way" to go on that is "the same" as how we have been going on, but we have lost our socially scaffolded confidence in our ability to discern what it is.

[17] Something very similar was proposed by Kuhn about scientific crises and the progressive resolution thereof (2012, pp. 161–162).

of security, and it is natural for me to be confident that we are– but this is no guarantee that at some point in the future I won't be "contradicted on all sides" (Wittgenstein, 2008, §614). This feature of the language-game picture of social disruption has implications for the project of anticipatory ethics. According to Philip Brey, "[t]he central problem for an ethics of emerging technologies is that we do not know the future, and therefore do not know which ethical issues will play out once the technology is fully developed and entrenched in society" (2012, p. 2). The language-game picture of social disruption compounds this issue of predictability: our inability to predict how the future will go extends to our future rule-following judgements as well as to future technological developments. Even if we could know in advance the "future devices, applications, uses and social consequences" (ibid.) of an emerging technology, the language-game picture denies us the possibility of ascertaining ahead of time whether we are "in a position" to respond to these things, to go on as before, "given our current resources," as Löhr puts it. This makes it very difficult to see how we could ever anticipate social disruption in advance.

Can the language-game picture make sense of the justificatory (as opposed to the merely descriptive or predictive) role played by appeals to rule-following in contexts of social disruption? Blackburn (1984, pp. 294–296) has objected to this picture of rule-following on the basis that the individual and the group considered as a whole are in exactly the same position when it comes to the justificatory status of their rule-following intuitions: the language-game picture simply replaces "whatever seems right *to me* is right" with "whatever seems right *to us* is right." But this objection can be avoided simply by recognising that the practice of justification only ever takes place– and only ever has a point– within a discourse: justifications *become such* through the process of demanding them from and offering them to one another (Kusch, 2006, p. 192). The very idea of a justification outside of the context of a discourse within which the justification is demanded and offered makes as little sense as the idea of it being 5 o'clock on the sun (Wittgenstein, 2009, §350). The idea of it being 5 o'clock only makes sense in certain surroundings; so too does the idea of justification. In both the case of the individual *and* of the group considered in isolation, as Blackburn correctly observes, seeming right *is* the only available criterion for being right, and therefore no question of justification can arise. But this is not so for the individual considered *in the context of the group to which they belong*. When the community agrees on the criteria for attributing justification, no appeal to consensus is necessary; rather, group consensus is what *makes it possible* for individuals to appeal to criteria for what counts as a successful justification. Therefore, the individual and the group are not in exactly the same position after all when it comes to the justificatory role of their rule-following intuitions: the language-game picture provides precisely that necessary element for making sense of a justified resolution of social disruption that the rules-as-rails picture was unable to provide.

Unlike rules-as-rails, the language-game picture therefore satisfies– albeit in a surprising way– the fourth desideratum proposed in Sect. 2: that of making sense of the notion of overcoming an episode of social disruption in a progress way. According to the language-game picture, social disruptions, when they do occur, resolve simply when deep agreement is re-established, and it becomes clear again what counts as going on in the same way.

At this point an important caveat is in order. There is no intrinsic fact about what counts as a going on in the same way in a disrupted context, such that resolution of social disruption can be pictured as our practices clicking back into place; rather, what counts as going on in the same way in such contexts is a *social fact* established by collective recognition, often retrospectively, once a de facto resolution has been reached. In this regard there is an element of moral luck in our collective ethical responses to SDTs (Williams, 1981; Nagel, 2012): if a proposal for responding to disruption is ultimately recognised as appropriate by the relevant community, then it counts as having been the right way to go on all along; if not, then it never was.[18]

To be sure, I am not advocating a kind of Pyrrhonism, or suspension of judgement, about what counts as going on in the same way or what situations or technologies count as socially disruptive. The practice of attributing the status of going on in the same way is essential to the liveability of human life, and giving up on that practice does not seem to be something of which human beings would be capable. Furthermore, I think we are perfectly entitled to continue to attribute the status of going on in the same way, precisely because of the brute fact that we generally agree about how to do so. We know social disruption when we see it, not because disruptive potential is out there to be seen independently of our perceptions of it, but because collectively regarding a given situation as disruptive is (part of) what makes it so. Instead of suspending judgement about what *really* counts as going on in the same way and what situations *really* count as socially disruptive, we ought to recognise the constitutive role played by our judgements in *making something count* as going on in the same way, and, correspondingly, in *making something count* as socially disruptive.

Viewed through the lens of the language-game picture of rule-following, paradigmatic examples of social disruption from history, including the emergence of paradigmatic SDTs like the mechanical ventilator, are moments in which deep agreement has broken down, but in ways that have hitherto been manageable and with which we have been able to cope, in the sense that deep agreement was ultimately reestablished after a period of controversy. However, in principle, deep agreement could break down catastrophically across the board. This might partly explain the profound disquiet many people feel about 21st century emerging technologies such as AI, quantum and nanotechnologies, and robotics: they are windows into the possibility of a whole world descending into what Durkheim called "anomie" (2005), or normlessness, threatening what Giddens calls an individual's "ontological security" (2003), their sense of the robustness, structure, and continuity of their life and self.[19]

---

[18] My use of "ultimately" here can be interpreted in two ways. One interpretation– the "Jamesean" interpretation– takes the individual disruptive episode as the relevant interval of time, and on this view the "ultimate" resolution is the one that suffices to close that particular disruptive episode (until the next one). The other– the "Piercean" interpretation– takes the relevant interval of time to extend indefinitely into the future. I am more inclined to the former interpretation but am not going to defend it here; Philip Kitcher defends something like the second interpretation in several places (e.g., 2014, pp. 246–249, 2021, pp. 36, 39).

[19] This connection between the language-game picture of social disruption and the thought of Durkheim and Giddens was pointed out to me by Nigel Pleasants, conversation with whom has greatly influenced the ideas expressed in this paper.

6 Conclusion

If we accept the language-game picture of rule-following it becomes quite mysterious how a socially disruptive phenomenon *brings about* the breakdown of deep agreement (the fifth desideratum from Sect. 2). Indeed, I'm sceptical about the possibility of anticipating social disruption in advance; having rejected the "rules-as-rails" picture of rule following, what ought to strike us as remarkable is not that consensus sometimes breaks down but that there is consensus about so much, most of the time, when there are no mutual states of grasping to explain our natural condition of deep agreement.

I think this is what Wittgenstein meant when he wrote that "[t]he real foundations of their inquiry do not strike people at all" (2009, §129). At bottom there is simply the fact that we do agree on what moral competence involves most of the time. When deep agreement does break down, we are unable to make progress in our inquiries, not because we need additional clues or hints to find the solution to a puzzle, but because of a genuine indeterminacy of criteria for identifying the right ways to move from perception to thought, thought to thought, and from thought to action. When we encounter a situation in which deep agreement is absent, we call this disruption. Rather than regarding the emergence of a socially disruptive technology as *bringing about* the breakdown of deep agreement, in a way that can be explained, predicted, and perhaps avoided or exploited, on the view I have proposed, socially disruptive technologies ought to be regarded simply as *occasions* for the breakdown of deep agreement, not its cause. A consequence of this is that social disruption is not something of which we can "get ahead."

I by no means regard what I have presented here as a fully fleshed-out picture of social disruption; my hope is to have demonstrated the relevance of the notion of rule-following for understanding social disruption, so that the picture can be fleshed out, improved, or possibly even rejected as inadequate, in the future. I hope to have drawn attention to some issues about how (or whether) occasions of social disruption can ever be satisfactorily explained, predicted, or progressively resolved, issues that only become visible once the role of rule-following in exercising the thinking techniques constitutive of inquiry, including conceptual engineering, has itself become visible. These issues, in my opinion, demand serious consideration, and I would welcome their further discussion.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

Anderson, E. (2014). *Social movements, experiments in living, and moral progress*. Case Studies from Britain's Abolition of Slavery'.

Anderson, E. (2015). Moral Bias and corrective practices: A pragmatist perspective. *Proceedings and Addresses of the American Philosophical Association*, *89*, 21–47.

Baker, R. (2013). *Before Bioethics: A History of American Medical Ethics from the Colonial Period to the Bioethics Revolution*. Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780199774111.001.0001

Baker, R. (2019). *The structure of moral revolutions: Studies of changes in the morality of abortion, death, and the bioethics revolution*. The MIT Press. (Basic bioethics).

Baker, G. P., & Hacker, P. M. S. (1984). On misunderstanding Wittgenstein: Kripke's private Language argument. *Synthese*, *58*(3), 407–450.

Blackburn, S. (1984). The individual strikes back. *Synthese*, *58*(3), 281–301.

Boorse, C. (1975). On the distinction between disease and illness. *Philosophy & Public Affairs*, *5*(1), 49–68.

Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, *44*(4), 542–573.

Bower, J. L., & Christensen, C. M. (1995). 'Disruptive technologies: catching the wave', *Harvard Business Review*, *73*(1), pp. 43–53. Available at: https://doi.org/10.1016/0024-6301(95)91075-1

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

Briggle, A. (2008). 'Real friends: how the Internet can foster friendship', *Ethics and Information Technology*, *10*(1), pp. 71–79. Available at: https://doi.org/10.1007/s10676-008-9160-z

Carel, H. (2007). 'Can I Be Ill and Happy?', *Philosophia*, *35*(2), pp. 95–110. Available at: https://doi.org/10.1007/s11406-007-9085-5

Carel, H. (2018). *Illness: The Cry of the Flesh*. 3rd edn. Routledge. Available at: https://doi.org/10.4324/9781315197999

Cherry, M. J. (1996). 'Bioethics and the Construction of Medical Reality', *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, *21*(4), pp. 357–373. Available at: https://doi.org/10.1093/jmp/21.4.357

Christensen, C. M. (1997). *The innovator's dilemma: When new technologies cause great firms to fail*. Harvard Business School Press. (The management of innovation and change series).

Cocking, D., & Matthews, S. (2000). 'Unreal Friends', *Ethics and Information Technology*, *2*(4), pp. 223–231. Available at: https://doi.org/10.1023/A:1011414704851

Danaher, J. (2017). 'Robotic Rape and Robotic Child Sexual Abuse: Should They be Criminalised?', *Criminal Law and Philosophy*, *11*(1), pp. 71–95. Available at: https://doi.org/10.1007/s11572-014-9362-x

Danaher, J. (2019). 'The Philosophical Case for Robot Friendship', *Journal of Posthuman Studies*, *3*(1), pp. 5–24. Available at: https://doi.org/10.5325/jpoststud.3.1.0005

Danaher, J., & Sætra, H. S. (2023). 'Mechanisms of Techno-Moral Change: A Taxonomy and Overview', *Ethical Theory and Moral Practice*, *26*(5), pp. 763–784. Available at: https://doi.org/10.1007/s10677-023-10397-x

Dodd, J. (2023). 'Rails Invisibly Laid to Infinity', *The Philosophical Quarterly*, *73*(1), pp. 84–104. Available at: https://doi.org/10.1093/pq/pqac012

Durkheim, E. (2005). *Suicide*. New York: Routledge. Available at: https://doi.org/10.4324/9780203994320

Engelhardt, H. T. (1975). 'The Concepts of Health and Disease', in H.T. Engelhardt and S.F. Spicker (Eds.) *Evaluation and Explanation in the Biomedical Sciences: Proceedings of the First Trans-Disciplinary Symposium on Philosophy and Medicine Held at Galveston, May 9–11, 1974*. Dordrecht: Springer Netherlands, pp. 125–141. Available at: https://doi.org/10.1007/978-94-010-1769-5_9

Fricker, M. (2007). *Epistemic Injustice*. Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Giddens, A. (2003). *Modernity and Self-identity: Self and society in the late modern age*. Stanford Univ. Press.

Goosens, W. K. (1980). 'Values, Health, and Medicine', *Philosophy of Science*. 2022/04/01 edn, *47*(1), 100–115. Available at: https://doi.org/10.1086/288912

Hermann, J. (2022). 'Socially Disruptive Technologies and Moral Certainty', in Eriksen, C. et al., *Philosophical Perspectives on Moral Certainty*. 1st edn. New York: Routledge, pp. 19–34. Available at: https://doi.org/10.4324/9781003178927-2

Hermann, J. (2025). 'Moral certainty, deep disagreement, and disruption', *Synthese*, *205*(3), p. 103. Available at: https://doi.org/10.1007/s11229-025-04948-1

Hopster, J. (2021). 'What are socially disruptive technologies?', *Technology in Society*, *67*, 101750. Available at: https://doi.org/10.1016/j.techsoc.2021.101750

Hopster, J. (2023). '6. Conceptual Disruption and the Ethics of Technology', in I. Van De Poel et al. (Eds.) *Ethics of Socially Disruptive Technologies*. 1st edn. Cambridge, UK: Open Book Publishers, pp. 141–162. Available at: https://doi.org/10.11647/obp.0366.06

Ihde, D. (1990). *Technology and the lifeworld: From garden to Earth*. Indiana University Press (The Indiana series in the philosophy of technology).

Kitcher, P. (1995). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. 1st edn. New York: Oxford University Press. Available at: https://doi.org/10.1093/0195096533.001.0001

Kitcher, P. (2014). *The ethical project*. Harvard Univ. Press.

Kitcher, P. (2021). *Moral Progress*. Edited by J.-C. Heilinger. Oxford University Press. Available at: https://doi.org/10.1093/oso/9780197549155.001.0001

Kripke, S. A. (1982). *Wittgenstein on rules And private Language: An elementary exposition*. Blackwell.

Kuhn, T. S. (2012). *The structure of scientific revolutions: 50th Anniversary Edition*. Fourth edition. Chicago: The University of Chicago Press.

Kusch, M. (2006). *A sceptical guide to meaning and rules: Defending Kripke's Wittgenstein*. Acumen.

Lakatos, I. (1970). 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (Eds.) *Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*. Cambridge: Cambridge University Press, pp. 91–196. Available at: https://doi.org/10.1017/CBO9781139171434.009

Lakatos, I. (1978). *The Methodology of Scientific Research Programmes: Philosophical Papers*. 1st edn. Edited by J. Worrall and G. Currie. Cambridge University Press. Available at: https://doi.org/10.1017/CBO9780511621123

Lane, B. (2023). 'Prospects for pure procedural moral progress', *Inquiry*, pp. 1–31. Available at: https://doi.org/10.1080/0020174X.2023.2220753

Lane, B. (2024).'Deep disagreement across moral revolutions', *Synthese*, *204*, 72. Available at: https://doi.org/10.1007/s11229-024-04710-z

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. *Shaping Technology/building Society: Studies in Sociotechnical Change*, *1*, 225–258.

Lennox, J. G. (1995). 'Health as an Objective Value', *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, *20*(5), pp. 499–511. Available at: https://doi.org/10.1093/jmp/20.5.499

Löhr, G. (2023a). 'Conceptual disruption and 21st century technologies: A framework', *Technology in Society*, *74*, 102327. Available at: https://doi.org/10.1016/j.techsoc.2023.102327

Löhr, G. (2023b). 'Do socially disruptive technologies really change our concepts or just our conceptions?', *Technology in Society*, *72*, 102160. Available at: https://doi.org/10.1016/j.techsoc.2022.102160

Marchiori, S., & Scharp, K. (2024). 'What is conceptual disruption?', *Ethics and Information Technology*, *26*(1), p. 18. Available at: https://doi.org/10.1007/s10676-024-09749-7

Margolis, J. (1976). 'The Concept of Disease', *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, *1*(3), pp. 238–255. Available at: https://doi.org/10.1093/jmp/1.3.238

Marx, K., & Engels, F. (2011). *The communist manifesto*. Penguin Books.

Maudlin, T. (2005). 'The Tale of Quantum Logic', in Y. Ben-Menahem (Ed.) *Hilary Putnam*. Cambridge: Cambridge University Press (Contemporary Philosophy in Focus), pp. 156–187. Available at: https://doi.org/10.1017/CBO9780511614187.006

McDowell, J. (1981). Non-cognitivism and rule-following. In S. H. Holtzman, & C. M. Leich (Eds.), *Wittgenstein: To follow A rule* (pp. 141–162). Routledge.

Moody-Adams, M. M. (1999). THE IDEA OF MORAL PROGRESS. *Metaphilosophy*, *30*(3), 168–185.

Nagel, T. (Ed.). (2012). 'Moral Luck', in *Mortal Questions*. Cambridge: Cambridge University Press (Canto Classics), pp. 24–38. Available at: https://doi.org/10.1017/CBO9781107341050.005

Nickel, P. J. (2020). 'Disruptive Innovation and Moral Uncertainty', *NanoEthics*, *14*(3), pp. 259–269. Available at: https://doi.org/10.1007/s11569-020-00375-3

Nickel, P. J., Kudina, O., & van de Poel, I. (2022). 'Moral Uncertainty in Technomoral Change: Bridging the Explanatory Gap', *Perspectives on Science*, *30*(2), 260–283. Available at: https://doi.org/10.1162/posc_a_00414

Pleasants, N. (2008). 'Institutional Wrongdoing and Moral Perception', *Journal of Social Philosophy*, *39*(1), 96–115. Available at: https://doi.org/10.1111/j.1467-9833.2007.00413.x

Popper, K. R. (2007). *Conjectures and refutations: The growth of scientific knowledge*. Routledge (Routledge Classics).

Putnam, H. (Ed.). (1979). 'A philosopher looks at quantum mechanics', in *Mathematics, Matter and Method: Philosophical Papers*. 2nd edn. Cambridge: Cambridge University Press, pp. 130–158. Available at: https://doi.org/10.1017/CBO9780511625268.009

Rehren, P., & Blunden, C. (2024). 'Let's not get ahead of ourselves: we have no idea if moral reasoning causes moral progress', *Philosophical Explorations*, 27(3), pp. 351–369. Available at: https://doi.org/10.1080/13869795.2024.2363876

Reznek, L. (1987). *The nature of disease*. London. (Philosophical issues in science).

Sætra, H. S. (2022). 'Loving robots changing love: Towards a practical deficiency-love', *Journal of Future Robot Life*, *3*(2), pp. 109–127. Available at: https://doi.org/10.3233/FRL-200023

Schumpeter, J. A. (2013). *Capitalism, Socialism and Democracy*. 5th edn. London: Routledge. Available at: https://doi.org/10.4324/9780203202050

Searle, J. R. (1995). *The construction of social reality*. Lane, The Penguin.

Swierstra, T., & Rip, A. (2007). 'Nano-ethics as NEST-ethics: Patterns of Moral Argumentation About New and Emerging Science and Technology', *NanoEthics*, *1*(1), pp. 3–20. Available at: https://doi.org/10.1007/s11569-007-0005-8

Vallor, S. (2012). 'Flourishing on facebook: virtue friendship & new social media', *Ethics and Information Technology*, *14*(3), 185–199. Available at: https://doi.org/10.1007/s10676-010-9262-2

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780190498511.001.0001

Veluwenkamp, H. (2022). 'Technology as Driver for Morally Motivated Conceptual Engineering', *Philosophy & Technology*, *35*(3), 71. Available at: https://doi.org/10.1007/s13347-022-00565-9

Verbeek, P. P. (2006). 'Materializing Morality: Design Ethics and Technological Mediation', *Science, Technology, & Human Values*, *31*(3), 361–380. Available at: https://doi.org/10.1177/0162243905285847

Verbeek, P. P. (2008). 'Obstetric Ultrasound and the Technological Mediation of Morality: A Postphenomenological Analysis', *Human Studies*, *31*(1), 11–26. Available at: https://doi.org/10.1007/s10746-007-9079-0

Williams, B. (Ed.). (1981). 'Moral luck', in *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press, pp. 20–39. Available at: https://doi.org/10.1017/CBO9781139165860.003

Winner, L. (1980). Do artifacts have politics? *Daedalus*, *109*(1), 121–136.

Winner, L. (2014). 'Technologies as Forms of Life', in R.L. Sandler (Ed.) *Ethics and Emerging Technologies*. London: Palgrave Macmillan UK, pp. 48–60. Available at: https://doi.org/10.1057/9781137349088_4

Wittgenstein, L. (2008). *On certainty*. Edited by D. Paul. Translated by G.E.M. Anscombe. Oxford: Blackwell.
Wittgenstein, L. (2009). *Philosophical investigations*. Rev. 4th ed. Edited by P.M.S. Hacker and J. Schulte. Translated by G.E.M. Anscombe. Chichester, West Sussex, U.K.: Wiley-Blackwell.
Wittgenstein, L. (Ed.). (1996). *Remarks on the foundations of mathematics*. Rev. ed., 4. print. Translated by G.H. von Wright. Cambridge, Mass.: MIT Press.