Synthese (2025) 205:103
https://doi.org/10.1007/s11229-025-04948-1

ORIGINAL RESEARCH

Moral certainty, deep disagreement, and disruption

Julia Hermann[1]

Received: 30 April 2024 / Accepted: 28 January 2025

Abstract

Wittgenstein's On Certainty has been a source of inspiration for philosophers concerned with the notion of deep disagreement (see Fogelin in Informal Logic 25(1):3–11, 2005; Pritchard in Topoi 40:1117–1125, 2021. https://doi.org/10.1007/s11245-018-9612-y). While Wittgenstein's examples of certainties do not include moral certainties, some philosophers have argued that an analogy can be drawn between certainty regarding the empirical world and moral certainty (Goodman in Metaphilosophy 13:138–148,1982; Hermann in On moral certainty, justification, and practice: A Wittgensteinian perspective, Palgrave Macmillan, Basingstoke, 2015; Pleasants in Inquiry 51(3):241–267, 2008. https://doi.org/10.1080/00201740802120673). Moral certainty manifests itself in our fundamental ways of feeling, thinking, and acting morally. It is closely related to an "agreement in form of life" (Wittgenstein in Philosophical investigations, Translated by G.E.M. Anscombe. Third ed., Basil Blackwell, Oxford, 1968, § 241.), which makes moral disagreements possible in the first place. In this paper, I aim to shed light on the phenomenon of moral deep disagreement by relating it not only to the notion of moral certainty but also to the concept of deep disruption as it is currently developed and discussed in the philosophy of technology. I argue that certainty, deep disagreement, and deep disruption are all located at the level of "bedrock practices" (Williams in Wittgenstein, mind and meaning, Routledge, Milton Park, 1999, p. 198), and that the fundamentality of their objects should be understood in terms of relationality and interconnectedness. Deep disagreements can occur through deep technology-induced disruption and can take the form of a disruption of deep conceptual agreement. Conceptual common ground can be re-established by continuous interaction and a collective process of moral articulation. Deep disruption and moral deep disagreement can lead to moral progress, for instance in the form of recognising and correcting an epistemic injustice.

Keywords Conceptual disruption · Deep disagreement · Moral certainty · Moral progress · Technology · Ludwig Wittgenstein

Extended author information available on the last page of the article

Published online: 21 February 2025

1 Introduction

Wittgenstein's *On Certainty* has been a source of inspiration for philosophers concerned with the notion of deep disagreement (see Fogelin, 2005; Pritchard, 2021). The hinge-metaphor that Wittgenstein in one place uses to refer to that which is certain (Wittgenstein, 1972, § 341) has inspired what is called "hinge epistemology" (see, e.g., Coliva & Doulas, 2022). While Wittgenstein's examples of certainties (or "hinges") don't include moral certainties, I and some other philosophers working in ethics and metaethics have argued that an analogy can be drawn between certainty regarding the empirical world and moral certainty (Goodman, 1982; Hermann, 2015; Pleasants, 2008). Moral certainty manifests itself in our fundamental ways of feeling, thinking, and acting morally (Hermann, 2015). Moral thinking and acting reveals an agreement in moral judgements. This agreement is a precondition for moral disagreement, which is not as pervasive as philosophers tend to assume. Moral deep disagreements seem to be best understood as disagreements involving conflicting moral certainties or hinges. Such deep disagreements are taken to be concerning because it seems impossible to overcome them.

In the philosophy and ethics of technology, a concept that is currently receiving much attention is that of disruption (Hopster, 2021; Löhr, 2023a and 2023b; Marchiori & Sharp, 2024; van de Poel et al., 2023). "Technosocial disruption" refers to the disruption of fundamental human practices, norms, values, and concepts by an interplay between technologies such as artificial intelligence, synthetic biology, gene-editing technologies, and climate-engineering technologies, and other components of the socio-technical systems they are part of, with technology playing a significant role. Technosocial disruption affects "deeply held beliefs, values, social norms, and basic human capacities", "basic human practices, fundamental concepts, [and] ontological distinctions" (Hopster, 2021, p. 6). It thus deeply affects us as human beings, our relationship to other parts of nature such as non-human animals and plants, and the societies we live in. For this reason, it is also referred to as "deep disruption" (ibid.). Among such deep disruptions are disruptions of fundamental concepts. When a deep conceptual disruption happens, our usual ways of applying fundamental concepts without much reflection are interrupted.

An example of this is the disruption of learning and teaching practices and fundamental beliefs, norms, and concepts related to them through the rapid development of generative AI. ChatGPT, for instance, not only disrupts routines related to assessment and grading, but also our understanding of what students should learn, i.e., the goals of education, and of the conditions in which a student's argument is their own. The disruption manifests itself in the fact that we cannot go on as before and are forced to rethink and possibly reorient our educational practices. One of the concepts that is being challenged in this case is that of learning.

In this paper, I aim to shed light on the phenomenon of moral deep disagreement by relating it not only to the notion of moral certainty but also to the concept of deep disruption. I argue that certainty, deep disagreement, and deep disruption are located at the same level– the level of "bedrock practices" (Williams, 1999, p. 198)– and that the fundamentality of their objects should be understood in terms of relationality and interconnectedness. My analysis of the relationship between the three concepts shows

that deep disagreements can occur through deep technology-induced disruption and can take the form of a disruption of deep conceptual agreement. Conceptual common ground can be re-established by continuous interaction and a collective process of "moral articulation" (Congdon, 2023). While I take this to hold for deep disagreements generally, I illustrate it by means of examples arguably involving *moral* deep disagreements. Deep disruption and moral deep disagreement, I argue moreover, can lead to moral progress, for instance in the form of recognising and overcoming an epistemic injustice.

A further aim of this paper is to show how the current research in the philosophy of technology on socially disruptive technologies can contribute to a richer understanding of the notion of deep disagreement, and how a Wittgensteinian understanding of certainty and deep disagreement can in turn contribute to our understanding of the phenomena of technology-induced social and conceptual disruption. Thereby, I also advocate for a recognition of the philosophy of the later Wittgenstein as providing rich resources for the philosophy of technology.

In Sect. 2, I present Wittgenstein's notion of certainty and the concept of moral certainty that philosophers inspired by Wittgenstein have developed. Then I turn to the notions of deep disagreement (Sect. 3) and deep disruption (Sect. 4). In Sect. 5, I explore the relationship between the concepts of moral and non-moral certainty, moral and non-moral deep disagreement, and deep disruption by looking at two examples: human-robot relationships and AI systems engaging in moral reasoning. I argue that in those examples, moral deep disagreements can occur through deep technology-induced disruption and can take the form of a disruption of deep conceptual agreement. Conceptual common ground can be re-established by continuous interaction and a collective process of moral articulation. In Sect. 6, I move on to argue that deep disagreements and disruptions can sometimes lead to moral progress, for instance in the form of correcting an epistemic injustice. In Sect. 7, I conclude.

2 Certainty

In the remarks published as *On Certainty*, Wittgenstein grapples with cases in which it seems impossible to reasonably doubt something, for instance the existence of one's hands[1], one's own name, or that the earth has existed long before one was born (Wittgenstein, 1972, §§ 4, 84, 91).[2] Some of his examples are taken from G.E. Moore, who claimed to know a number of things with certainty and attempted to prove the existence of the external world via the indubitability of the existence of his hands (Moore, 1959a,1959b). In the cases contemplated by Wittgenstein, we have no

[1] It is under *normal circumstances* that the existence of my hands is certain. We can imagine exceptional circumstances, in which doubt was possible, for instance the situation where I had an accident and woke up after an operation, not knowing whether my hands had been amputated.

[2] *On Certainty* is comprised of notes written by Wittgenstein during the last eighteen months of his life. It is taken from four different notebooks and a bunch of separate sheets (van Gennip, 2008, p. 52). Wittgenstein did not have time to revise these notes, in which he attempts to get to grips with something by turning around the same questions and examples time and again. At some points, he expresses his dissatisfaction with the way in which he formulated his thoughts.

idea what a doubt would look like (Wittgenstein, 1972, § 32; see also Wittgenstein, 1992, p. 46). How, for instance, can I doubt that the things I am now typing these words with are my fingers? The indubitability at stake here is neither conceptual nor psychological. We could call it "logical" in the sense that it is excluded by the logic of the language game. In one of his later remarks, Wittgenstein (1972, § 674) states that he "can enumerate various typical cases, but not give any common characteristic". The heterogeneity of Wittgenstein's examples should not be underestimated. They range from things that are certain for most human beings (e.g., "I am a human being"; ibid., § 4) to what some have called "personal certainties" (e.g., Moyal-Sharrock, 2004, p. 101), such as "My name is L.W." (Wittgenstein, 1972, § 425). While it seems to hold for some of the examples that they "underlie all questions and all thinking" (Wittgenstein, 1972, § 415), others can rather be said to be presupposed by more specific languagegames.

Something is certain not in virtue of some intrinsic feature, but in virtue of how it is treated by a community of speakers, which may extend to the whole human form of life. Wittgenstein makes use of a rich reservoir of metaphors to describe the phenomenon of certainty. Some of these metaphors are foundationalist (e.g., "bedrock" (ibid., § 498), "rock bottom of my convictions" (ibid., § 248), "foundations of all judging" (ibid., § 614)), while others are coherentist (e.g., "[what] stands fast […] is […] held fast by what lies around it" (ibid., § 144), "the movement around [the axis] determines its immobility" (ibid., § 152), the "foundation-walls are carried by the whole house" (ibid., § 248)). Peter Winch (1998, p. 198) has emphasised the fittingness of the metaphor of the axis, as opposed to the hinge metaphor. This is the passage in which Wittgenstein uses the hinge-metaphor: "That is to say, the questions that we raise and our doubts depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn" (Wittgenstein, 1972, § 341). Unlike a hinge, which is fixed and the existence of which is independent of the motion of the door, an axis is determined by the movement around it. It has "no existence or meaning apart from the movement" (Winch, 1998, p. 198). For instance, the certainty that I am a human being is determined by the movement around it, which includes empirical statements about someone (e.g., a fictional character) or something (e.g., an embryo) being a human being, which can be true or false. Although that which is certain plays a foundational role in our practices of judging, asserting, questioning, justifying, and so forth, it should not be conceived of as a solid and immoveable foundation. It is the practice of judging that turns something into a foundation (Krebs 2007, p. 108). That practice is embedded in a form of life.

Wittgenstein suggests saying about the things Moore claims to know that they "stand fast" for him, and indeed for everyone (Wittgenstein, 1972, § 152).[3] Treating them as fixed, as beyond doubt, is characteristic of human practices. That which is certain is neither true nor false. It belongs to the background against which we distinguish between true and false. It is true only insofar as it is an "unmoving foundation of [our] language-games" (ibid., § 403). Propositions such as "I am a human being"

[3] Since the formulations "stand fast" and "held fast" are common in the literature on *On Certainty*, I also use them, along with "are fixed", "are certain" and "are beyond doubt" as synonyms for "stand fast". "Are fixed" should be read as "are being treated as fixed".

or "the earth existed long before I was born" have no use in the languagegame but play a role similar to that of a grammatical proposition. They belong to the description of a languagegame, and thus to logic (see Wittgenstein, 1968, §§ 251 and 496; Wittgenstein 1972, § 628). The Wittgensteinian logic is related to humans. It expresses that which we in our contingent world treat as necessary, not that which is absolutely necessary independently of humans (see Moyal-Sharrock, 2004, pp. 155 f.). It is crucial for Wittgenstein's account of certainty that doubt is the exception (Wittgenstein, 1993, p. 106) and presupposes certainty (ibid., p. 115). He illustrates the claim that doubt is an exception by means of an example of non-linguistic behaviour (ibid., p. 108). He imagines a mother who reacts to her crying child holding its cheek by going through the motions of sceptical behaviour with regard to the pain of her child, making no attempts to console it, and only exceptionally behaving as if she believed her child that it is in pain. We would not call that kind of behaviour scepticism or doubt. To illustrate the second point– that doubt presupposes certainty– Wittgenstein refers to a child who first has to take numerous things for granted before it can meaningfully doubt anything (Wittgenstein, 1972, § 160 and 283; see also § 329 and 522).

Wittgenstein also says that the propositions that stand fast describe someone's world-picture.[4] The term "world-picture" highlights, amongst other things, the fact that certainties are not isolated from one another but form an organic unity. We do not acquire them in isolation. My world-picture is "the inherited background against which I distinguish between true and false" (Wittgenstein, 1972, § 94). That which is certain is "anchored in all my *questions and answers*, so anchored that I cannot touch it" (ibid., § 103). The certainties that Moore lists and that are often referred to as "Moorean propositions" are believed by all "normal", mentally sane people.[5]

Wittgenstein moreover emphasises that he is not concerned with subjective certainty, i.e., a feeling of certainty: "one isn't trying to express even the greatest subjective certainty, but rather that certain propositions seem to underlie all questions and all thinking" (Wittgenstein, 1972, § 415). The phenomenon he is trying to get to grips with is objective in the sense that a mistake is logically impossible, i.e., excluded by the rules of the languagegame (ibid., § 194).

According to Moyal-Sharrock's non-propositional reading of *On Certainty*, which I largely agree with, "objective certainty" is the certainty with which we act (Moyal-Sharrock, 2004, p. 72).[6] It is "enacted" (ibid.). Our actions show, e.g., that it is certain for us that we are human beings. Nigel Pleasants (2009, p. 670) emphasises that "[w]hat underlies our epistemic practices and capacities is not itself an epistemic practice or capacity, but our fundamental ways of being and acting in the world". A crucial passage for the non-propositional reading of *On Certainty* is the following:

[4] I follow Wittgenstein's translators in using "world-picture" instead of "world-view", although the former term does not convey the full meaning of the German term "Weltbild", which seems to be better captured by "world-view".

[5] I admit that the term "normal" is problematic. I want to emphasise that "normal" here does not refer to neurotypical as opposed to neurodivergent people. Wittgenstein uses the term as including anyone who is not insane in the sense of speaking and acting in an utterly incomprehensible way.

[6] Other interpreters defend a propositional reading (e.g., Williams, 2004). Wittgenstein does not use the term "objective certainty", but emphasises, as was noted above, that he is not concerned with subjective certainty (Wittgenstein, 1972, § 415).

> Giving grounds, however, justifying the evidence, comes to an end; - but the end is not certain propositions' striking us immediately as true, i.e. it is not a kind of *seeing* on our part; it is our *acting*, which lies at the bottom of the language-game.
>
> *On Certainty*, § 204[7]

What does "our acting" refer to? As I have argued in previous work (Hermann, 2015, p. 58), it refers to a combination of "shared, natural reactions, unreflective uses of language, unreflective non-linguistic actions, and the regularity with which we act". This regularity is also referred to with the notion "agreement in judgements" (Wittgenstein, 1968, § 242). Agreement in judgements refers to shared ways of acting and judging. Meredith Williams speaks about "bedrock practices", which are constituted by "the harmonious 'blind' agreement in words and deeds of a group of people over a period of time", where "blind" means that it does not result from the self-conscious and explicit application of rules (Williams, 1999, p. 198). As Wittgenstein tries to show in the *Philosophical Investigations*, human language presupposes agreement ("Übereinstimmung")[8] in judgements (Wittgenstein, 1968, § 242). The notion of certainty is thus closely related to that of agreement in judgements.

While Wittgenstein's examples of certainties do not include *moral* certainties, I and some other philosophers working in ethics and metaethics have argued that an analogy can be drawn between certainty regarding the empirical world and moral certainty (Goodman, 1982; Hermann, 2015; Pleasants, 2008). Just as the way we normally use our hands and talk about them reveals our certainty regarding their existence, our moral conversations and practices (e.g., the ways in which we react to people who need help) reveal the certainty that helping people in need is right, harming people is wrong, and so forth. Practices of moral deliberation and justification presuppose shared ways of reacting to certain behaviour, situations, and interactions. They also presuppose some agreement on the application conditions for moral concepts. There could be no disagreement about the wrongfulness of a particular act of killing, e.g., the killing of an animal for the purpose of medical research, in the absence of agreement on the wrongness of paradigmatic cases of killing (see Pleasants, 2009). Moral certainty manifests itself in our fundamental ways of feeling, thinking, and acting morally, which more sophisticated moral practices such as justifying and deliberating presuppose (Hermann, 2015). Moral thinking and acting reveals an agreement in moral judgements, which is a precondition for moral disagreement. As I have argued (Hermann, 2015), moral certainty depends on moral competence. For morally competent agents, the prima facie wrongness of killing, harming, and stealing is certain. These certainties are "the axes around which our moral doubts and justifications turn" (ibid., p. 93).

[7] See also § 110: "[...] As if giving grounds did not come to an end sometime. But the end is not an ungrounded presupposition: it is an ungrounded way of acting."

[8] The German term "Übereinstimmung" is not well translated by the term "agreement", since the latter suggests that people have come to agree on these judgements. What Wittgenstein means is that people happen to hold the same judgements, or to judge in the same way.

3 Moral and non-moral deep disagreement

On Certainty has been a source of inspiration for philosophers concerned with the notion of deep disagreement (see Fogelin, 2005; Pritchard, 2021). After briefly presenting the most prominent Wittgenstein-inspired accounts of deep disagreement, I will focus on the account of Robert Fogelin and elaborate on it. According to Pritchard (2021, p. 1117), deep disagreements concern "the most basic commitments on the parts of the disagreeing subjects". Pritchard (ibid.) further specifies those commitments as "the kinds of commitments that reflect our ways of seeing the world, at a most fundamental level, and in doing so often reflect our deeply held values". He mentions disagreements between religious and non-religious people as obvious examples of deep disagreements. He also refers to "basic political or ethical disagreements" as examples (ibid., p. 1118). While Fogelin (2005, p. 11) takes deep disagreements to be "by their nature, not subject to rational resolution", Pritchard (2021, p. 1119) argues that at least in the class of deep disagreements that he is mainly interested in they are rationally resolvable (ibid.). Ranalli (2021) proposes an understanding of deep disagreements that does not make rational irresolvability a necessary characteristic of such a disagreement.

In the terminology of "hinge epistemology", some deep disagreements are of the kind that the parties disagreeing have different hinge commitments. Ranalli (2021, p. 986) presents what he calls the "Wittgensteinian theory of deep disagreement" thus: "Deep disagreements are disagreements over hinge commitments." He characterises hinge commitments as "the background presuppositions of our world views and general areas of inquiry, such as physics, history, or geology" (ibid.). Pritchard (2021, p. 1118) writes: "Given the nature of our hinge commitments, it seems almost definitional that a clash of hinge commitments is a deep disagreement."

The examples of moral deep disagreements provided by Ranalli include disagreements over whether persons have moral status or whether murder is wrong. These examples are problematic though. Persons by definition have moral status; murder is by definition morally wrong. Were two parties to disagree about those things, that would suggest that one of them lacks moral competence (see Hermann, 2015). By contrast, if the first disagreement were about the moral status of a humanoid robot with humanlike capacities, it would be conceivable as a disagreement between two morally competent agents.

On Fogelin's characterisation of deep disagreements, "they persist even when normal criticisms have been answered" and are "immune to appeals to facts" (Fogelin, 2005, p. 8). The first example he mentions is that of disagreement about the morality of abortion, which concerns the moral status of the fetus (ibid.). Fogelin (ibid., p. 9) points out that the source of a deep disagreement does not consist of isolated propositions but lies in "a whole system of mutually supporting propositions (and paradigms, models, styles of acting and thinking) that constitute […] a form of life". He approaches the issue of deep disagreement via the notion of argument. He argues that arguments require a "normal" argumentative context, i.e., a context in which there is common ground– "a shared background of beliefs and preferences"– to refer to (Fogelin, 2005, p. 7). In contexts that are not normal in this sense, the "conditions for argument" are absent. According to Fogelin, such non-normal contexts are situ-

ations in which there is "a clash of framework propositions", which generates the disagreement (ibid., p. 8). This might not be the only form though that non-normal contexts can take. As I argue below (Sect. 5), another form is a conceptual disruption.

"Normal", i.e., not deep, disagreements manifest themselves in arguments. In a Wittgensteinian spirit, Fogelin conceives of arguing as something that people do. It is an activity that can have a variety of purposes. Arguing involves providing reasons in support of one's own position and to undermine the position of the other party. Given that the disagreeing parties share many beliefs and preferences (their common ground), it is possible to convince the other and to be convinced by them. Thus, it is possible to rationally resolve the argument. Fogelin conceptualises deep disagreement as a disagreement in the absence of the conditions for genuine argument. While "[t]he language of argument may persist, […] it becomes pointless" (ibid., p. 7).

This phenomenon can be illuminated further by drawing on Wittgenstein's notion of "agreement in judgements" (Wittgenstein, 1968, § 242). An agreement in judgements is not an agreement that is the result of an argument. It is not the kind of agreement that can be reached by exchanging reasons with the aim of convincing one another. Agreement in judgements refers to shared ways of acting and judging. This includes shared ways of using words. Wittgenstein (1968, § 241) also calls this "agreement in form of life". In *Remarks on the Foundations of Mathematics* (VI 39), he writes that "the phenomenon of language is based on regularity, on agreement in action". As I have argued elsewhere (Hermann, 2015), certainty is closely connected to this agreement in judgements, action, or forms of life. In a similar vein, Williams (1999, p. 198) argues that our language games of justification and knowledge presuppose "conformity within bedrock practices". In cases of deep disagreement, I suggest, the agreement in judgements that underlies and enables all acts of deliberation and argumentation has become disrupted. This can happen in different ways. For instance, it can be the result of an encounter between members of very different cultures. Fogelin would probably see this as a case in which there is a "clash of framework propositions". I prefer to formulate it in non-propositional terms. In such a case, the people involved do not act and judge in concert at the level of bedrock. Another way in which this can happen is through the introduction of new technologies. In the examples discussed below (Sect. 5), new technologies disrupt the unreflective ways of applying concepts such as "moral agency" and "responsibility", leading to disagreement about the application conditions of those concepts. It is the phenomenon of technology-induced disruption to which I will now turn.

4 Deep disruption

Recent scholarship in the philosophy and ethics of technology is investigating a phenomenon that has been coined "technosocial disruption" (Hopster, 2021, p. 6). It is particularly concerned with such disruption at a fundamental level– the level that deep disagreement is also located at. It deeply affects us as human beings, our relationship to other parts of nature such as non-human animals and plants, and the societies we live in. According to Hopster, deep disruption brings about different kinds of

uncertainty, including "conceptual ambiguity and contestation, moral confusion, and moral disagreement" (ibid., p. 7).

How should we understand this fundamentality? Among the things affected by technosocial disruption are fundamental concepts, i.e., concepts that play a fundamental role in human life. Löhr (2023b) links the fundamental nature of a concept to its inferential connectedness to other concepts. A concept is fundamental if changes of it will affect many other concepts (Löhr, 2023b: 5). Löhr (2023b, p. 5) claims that "the more fundamental a concept is, the more will a disruption of [its] use affect other concepts that are inferentially related to it, especially given the important mediating and creative role of concepts". A fundamental concept thus plays a central role in human practices and is intricately related to several other concepts. As noted above, in *On Certainty*, Wittgenstein uses both foundationalist and coherentist metaphors to describe the phenomenon of certainty. As I have argued in previous work (Hermann, 2015, p. 57), Wittgenstein should not be interpreted as understanding *justification* in terms of coherence. In *On Certainty*, considerations of coherence can explain the fundamental role that some beliefs seem to play (ibid.). In the same vein, Löhr explains the fundamental role of some concepts by reference to their relations to other concepts. Fundamentality is thus understood in terms of relationality and interconnectedness.

As mentioned above, when reflecting on the phenomenon of certainty, Wittgenstein also uses the term "world-picture". Lane (2024, p. 8) characterises a world-picture as a "concept-infused representation of reality". Löhr (2022, p. 2) speculates that "new and future technologies will likely further disrupt not just our view of the world but the concepts that allow us to access the world in the first place". This shows the close connection between certainty and some kinds of concepts and makes it interesting to analyse the relationship between certainty and conceptual disruption. Perhaps conceptual disruptions can be understood as disruptions of world-pictures via the disruption of the concepts that constitute those world-pictures.[9]

Löhr (2022, p. 4) defines a conceptual disruption as "[a]ny intentional or unintentional challenge or interruption of the ways in which the individual or group has intuitively classified individuals, properties, actions, situations, or events, leading to *classificatory uncertainty*, i.e., uncertainty about the application conditions of a word or concept". Marchiori and Scharp (2024, p. 4) propose a slightly different definition: "a *conceptual disruption* is an interruption in the normal functioning of a concept, cluster of concepts, or conceptual scheme". They object to Löhr's definition that classification is a function only performed by predicates and that the definition thus does not apply, for instance, to logical connectives (ibid., p. 3). I adopt Marchiori and Scharp's definition, as I take the interruption of the normal functioning of a concept to be the core of conceptual disruption. Hopster and Löhr (2023) distinguish three types of conceptual disruption: conceptual gaps, conceptual overlaps, and conceptual misalignments.[10] Conceptual gaps occur "if the new artifacts, states, or events generated by new technologies do not fit any single familiar category" (Löhr, 2023b, p. 6).

[9] In the examples commonly discussed in the literature on deep disagreement, e.g., Young Earth Creationists, a difference in worldview leads to deep disagreement (Lagewaard, 2021, p. 1589).

[10] Marchiori & Scharp, 2024 propose a different typology.

A conceptual overlap occurs if there seems to be an equally good fit between the new artifacts, states, actions, or events and at least two familiar concepts (ibid., p. 6). In the case of a conceptual misalignment, concepts are in tension with norms and values (Marchiori & Scharp, 2024, p. 17).

The way in which Löhr describes the absence of conceptual disruption, i.e., the normal functioning of a concept, echoes how Wittgenstein writes about human action that manifests certainty. Löhr (2023b, p. 6) writes that in the absence of a conceptual disruption, we apply our concepts "normally without much reflection". He (ibid.) describes one's concepts as being "in the background". A concept is said to be disrupted in the sense that "we can no longer simply apply it without deliberation" (ibid.). Löhr (ibid.) is implicitly referring to our bedrock practices when he writes that a conceptual disruption "interrupts our unreflective or tacit classification in a way that usually demands reflection and reasoning". Explicitly referring to Wittgenstein's *Philosophical Investigations*, Löhr writes that we cannot "go on" as before. In the case of a conceptual disruption, our "customary, intersubjective ways of acting with words" (Tully, 2003, p. 26), which manifest what we could call "conceptual certainty", are being disrupted. If a conceptual disruption takes place at the level of fundamental concepts, this destabilises the order that we have imposed on the world with our conceptual tools, turning organised information into "blooming, buzzing confusion" (James, 1981, p. 462; see Stroud, 2019, p. 24).

5 Moral certainty, disagreement, and disruption

The previous section shed light on the connection between the phenomenon of conceptual disruption and the agreement in judgements that characterises our bedrock practices. In the following, I will connect these phenomena to moral and non-moral deep disagreement. Hopster (2021, p. 7) relates technosocial disruption to uncertainty and moral disagreement, arguing that the disruption brings about different kinds of uncertainty, including "conceptual ambiguity and contestation, moral confusion, and moral disagreement". In the literature on uncertainty in the context of emerging technologies, uncertainty is not understood as a counterpart to certainty in the Wittgensteinian sense. Nickel (2019, p. 4), for instance, reflects upon the "uncertainty-creating characteristics of modern data practices". He defines "practical uncertainties" as "things we *do not know* and have an *interest* in knowing" (ibid., p. 2). If people do not know what will happen with their data, they cannot know their own obligations and do not have the usual reasons to trust others. Thus Nickel, unlike Wittgenstein, links certainty to knowledge. The uncertainty in this case is due to a lack of ordinary knowledge, not to a lack of certainty as Wittgenstein conceives of it. However, we can also interpret the situation as one in which certainties in Wittgenstein's sense of the term are being disrupted. That would include the certainty that much personal information about oneself is unknown to others. Modern data practices disrupt "our trust in the world" (Moyal-Sharrock 2004, p. 199), which manifests certainty. This trust in the world is not based on reasons; it is the trust that is simply there in the absence of reasons to be skeptical. This Wittgensteinian perspective

on modern data practices brings to light a disruption that goes deeper than the one described by Nickel.

Hopster likely understands uncertainty similar to Nickel and others, but interestingly he classifies moral disagreement as a kind of uncertainty. What kind of moral disagreement is this? When thinking of conceptual disruption as discussed above, the disagreement is likely to concern the application conditions of moral concepts. For instance, disagreement about whether humanoid robots are moral agents, have human rights, and deserve respect. Such a disagreement can be characterised as deep because it concerns the application conditions of *fundamental* concepts, such as moral agency. If the use of that concept is disrupted, this will affect many other concepts to which it is inferentially related, such as "moral responsibility", "autonomy", "moral rights", "moral duties", "respect", and so forth.[11] However, an understanding of moral deep disagreement as, following Fogelin (2005) and Pritchard (2021), a disagreement involving conflicting moral certainties or hinges and thus as a disagreement between people with fundamentally different moral world-pictures, does not fit this case. In situations of conceptual disruption, it is not an adequate description. Here the disagreement is among people whose assumably similar moral world-pictures have been disrupted with the effect that they disagree about how to apply a whole cluster of moral concepts. The disagreement could also be about how to fill a conceptual gap., i.e., about whether one or more concepts in a cluster need to be revised or new concepts introduced. Just as in the cases discussed by Fogelin, though, the conditions for argument are absent. To further analyse the relationship between certainty, deep disruption, and moral deep disagreement let us look at two examples.

First, consider the possibility of human-robot friendship and love, a topic that has received considerable attention in the philosophical literature (see, e.g., Nyholm & Frank, 2017). Should we take people who claim to be in love with a robot seriously? Are social, humanoid robots, at the current level of development, the kinds of entities that can be our friends and partners in romantic relationships? As the discussion about this shows, the development of such robots is challenging a whole cluster of concepts, including "love", "intimate relationship", "friendship", "reciprocal relationship", "emotional attachment", "trust", "moral obligations", and so forth. How should we classify the close relationships that some people have or claim to have with robots? Do concepts such as "intimacy", "friendship" and "love" apply to them? Should we subsume them under our existing concepts, or should we come up with revised or novel concepts that can capture the experiences of those reporting close relationships with robots? These questions suggest that we are faced with a conceptual gap. Conceptual gaps involve disagreement about how to classify something, which comes with uncertainty regarding what is morally required. Prior to the con-

[11] The conceptual cluster in which these concepts are closely related is not universal but historically and theoretically situated. It developed out of the Enlightenment period and is greatly influenced by Immanuel Kant. In a context in which these concepts play different roles and have different meanings, the disruptions would be different. In African contexts, for instance, personhood is closely related to the community, and agency is a relational phenomenon (see, e.g., Metz, 2007). The central value of interdependence stands in stark contrast with the Kantian autonomous, rational agent. I thank Dominic Lenzi for encouraging me to make this theory-dependence of the conceptual cluster explicit.

ceptual disruption, there was a "deep agreement"[12] underlying the application conditions of the above-mentioned concepts. It was certain that they apply to relationships between humans, and that being in such a relationship requires an ability to reciprocate on both ends. While it varies among groups and individuals what kinds of beings apart from humans are taken to be capable of being friends or lovers– e.g., some people believe that some sorts of animals can be our friends– there is deep agreement that humans can only be friends with sentient beings.[13]

Another example is the rapid development of artificially intelligent systems. ChatGPT impresses people with its ability to write, upon being given some prompts, recipes, poems, philosophical essays, and presentations. It can be used for a variety of purposes, including getting inspiration for new research topics, building a legal or moral argument, and getting recommendations. The better it gets at doing all those things that we used to think only human beings could do (reasoning, arguing, composing, etc.), the more tempting it becomes to delegate not only complex calculations and analyses to those systems, but also moral decision-making. Imagine a case in which the task of deciding who out of a group of patients receives a scarce life-saving treatment is handed over to such a system. This raises several pressing questions. Who or what is morally responsible for the decision? Under what conditions is the decision morally justified? Does the concept of moral responsibility apply to such systems? And do the related concepts of moral agency, intentionality, and autonomy apply to them? Can such systems be objects of justified blame? The use of the technology for such purposes thus leads to uncertainty regarding the use and applicability of several fundamental concepts. With that comes uncertainty regarding whom/what to hold responsible, whom/what to trust, and regarding one's own moral obligations and responsibilities. Prior to the conceptual disruption, there was a deep agreement underlying responsibility ascriptions. It was certain that only human agents can make moral decisions and carry responsibility for them. This deep agreement underlay all disagreements about who was responsible in concrete cases.

What disagreements can we envision with respect to these examples? Regarding the example of human-robot friendship and love, disagreement concerns the applicability of concepts such as "intimacy", "friendship", and "love" to such relationships. There is thus disagreement about how to classify these relationships, which comes with uncertainty regarding what is morally required. Regarding the example of AI systems making moral decisions, we can envisage disagreement about whether such systems can be morally responsible, i.e., whether the concept "moral responsibility" applies to them. This disagreement qualifies as deep in the sense that it concerns the application conditions of a fundamental concept. The concept of moral responsibility is closely related to concepts such as "moral agency", "autonomy", "moral rights", "moral duties", "respect", and so forth. It is also deep in the sense that it concerns people's most basic commitments (Pritchard, 2021, p. 1117), in this case commit-

[12] I owe this term to Benedict Lane.

[13] Young children who claim to be friends with a stuffed animal form an exception to this. Due to their lively imagination, they experience their favourite teddy bear as someone who listens to them and is there for them unconditionally. Given that these "friendships" depend on the children experiencing the teddy bear as a sentient being, thanks to their imagination, they do not provide a counterargument to the point made.

ments concerning who and what to hold morally responsible, whom and what to trust and whom and what to ascribe moral agency to.

What can we learn from these deep disagreements brought about by technology-induced conceptual disruption for the moral deep disagreements philosophers have been worried about (for example, disagreement about the morality of abortion)? Moral deep disagreements are taken to be concerning because it seems impossible to overcome them. Reflection upon cases of technology-induced conceptual disruption suggests a way in which deep disagreement could disappear. The common conceptual ground that is lacking in the case of the deep disagreements brought about by technology-induced conceptual disruption can be created by conceptual change of the form of concept revision, concept elimination, and the introduction of novel concepts. The development of new concepts involves creativity and imagination (Congdon, 2023).

Might this also be a way for dealing with disagreement about, for instance, the morality of abortion? In the abortion case as described by Fogelin, the parties disagree about the moral status of the fetus. They thus disagree, amongst other things, about the application conditions of the concepts "moral status", "personhood", etc. Given that the source of a deep disagreement lies in "a whole system of mutually supporting propositions (and paradigms, models, styles of acting and thinking) that constitute […] a form of life" (Fogelin, 2005, p. 9), coming to a point of agreement is not easy.

To approach the question as to how common ground could be established, let us take a look at how deep agreement comes about according to Wittgenstein. Agreement in judgements seems largely the result of socialisation. In his remarks on rule-following, Wittgenstein (1968, § 198) writes with regard to the connection between the expression of a rule such as a sign-post and his actions: "I have been trained to react to this sign in a particular way, and now I do so react to it." He subsequently denies that the connection is merely causal, pointing out that his formulation "indicated that a person goes by a sign-post only in so far as there exists a regular use of sign-posts, a custom". People who have been brought up in a similar environment, participating in the same practices, are in tacit agreement regarding a whole lot of things. The source of deep agreement suggests that restoring deep agreement after a deep technology-induced disruption requires participating in shared practices involving the new technology. I would assume, for instance, that by interacting with social robots in the context of shared practices, agreement as to how to classify these robots emerges over time. This implies that the possibilities for agents to intentionally create common ground are limited. Steps that can be taken include facilitating continuous interaction and, arguably, engaging in the critical examination and revision of relevant concepts. These steps presuppose that some agreement is already in place. That thin agreement provides the basis for reaching a more robust deep agreement.

It follows from the above that interaction is key if we consider how opponents in the abortion debate could reach the point where they can be said to be in deep agreement. Ideally, there would be interaction between religious people, non-religious people, pregnant people and their partners, doctors, philosophers, ethicists, and so forth. Here I am not thinking of debates between all those "stakeholders", at least not primarily. Rather, I envisage informal conversations and joint activities (e.g.,

creative workshops in which members of the different stakeholder communities visualise their understanding of a foetus or engage in role-playing or non-verbal embodiment exercises). This leads me to the issue of moral progress.

6 Disagreement, disruption, and moral progress

As we have seen, processes of technosocial disruption can give rise to deep disagreements about how to understand and apply moral concepts, or about whether a concept should be revised, eliminated, or a novel concept be introduced. These disruptions and disagreements can lead to moral progress. How?

A way in which a conceptual disruption can contribute to moral progress is by paving the way for correcting an epistemic injustice (see Hopster 2024). Deep conceptual disruptions can prompt what Congdon (2023, p. 3), inspired by Charles Taylor, calls "moral articulation". Some instances of moral articulation, such as the creation of the concept "sexual harassment", constitute moral progress. In many cases, the progressive character of a conceptual innovation lies in its ability to help addressing and correcting an injustice.

Congdon (2023, p. 51) describes moral articulation as "the dynamic activity of forming new words and conceptual schemes in order to bring previously inchoate, unprecedented, or marginalized forms of moral experience to expression, in ways that allow the sharing of moral meanings for a broader collective". Moral articulation is a creative process, in which emotions play a central role, and which happens both at the individual and at the social level. It starts with the experience of "discursive breakdown", which is an experience of the limits of language (ibid., p. 18). In situations of discursive breakdown, one experiences the meaning of something—an event, an action, a relationship—as elusive. The inability to describe the moral quality of a particular kind of experience with the existing conceptual repertoire prompts a process of articulation that, if successful, culminates in the development of a novel concept that transforms those experiences. One of Congdon's main examples is the development of the concept "sexual harassment".[14] The concept enables the expression of certain forms of experiences as instances of a particular type of injustice. "Through concepts", writes Congdon (ibid., p. 59), "we see the world in ethical colours". The concept of sexual harassment is the fruit of a collective struggle to articulate the specific moral quality of types of behaviour that had previously been obscure. While "sexual harassment" refers to "real, morally significant forms of suffering" (ibid., p. 51), the process of moral articulation involves a transformation of those experiences. It transforms an experience of suffering that involved a diffuse feeling of being wronged into an experience of being wronged in a specific way that forms part of a pattern and deserves moral blame and legal sanctioning. The "mastery of a language of sexual harassments" enables people to perceive "instances of unwelcome sexual attention"– a comment about one's looks, a joke, a touch—as "manifesting a single, unified, and purposive action-type" (ibid., p. 61). Importantly, this concept constitutes not simply an addition to the existing conceptual repertoire but

[14] This example also figures prominently in Fricker, 2007.

part of a broader transformation of our conceptual schemes (ibid., p. 50). According to Congdon (ibid., p. 222), the process in which the concept "sexual harassment" has been developed and adopted by a growing number of people constitutes an instance of moral progress.

Lagewaard (2021) explores the previously unnoticed connection between deep disagreement and epistemic injustice. She argues that "sometimes ordinary disagreements become deep as a result of epistemic injustice" (ibid., p. 1572). She calls such a case of deep disagreement "injustice-based deep disagreement" (ibid., p. 1580). In the context of this paper, it is particularly interesting that Lagewaard shows by means of the example of racism in the Netherlands how a lack of shared concepts can lead to hermeneutical injustice. Hermeneutical injustice is one of the two types of epistemic injustice distinguished by Miranda Fricker, the other being testimonial injustice. "Hermeneutical injustice" refers to cases in which, due to identity prejudices, a person who belongs to a group that is kept from participating equally in the "generation of social meanings" is thereby "put at an unfair disadvantage when it comes to making sense of a significant area of their social experience" (Fricker, 2013, p. 1319).[15]

As I argued above, shared concepts are part of the common ground that "normal" disagreements presuppose. If a lack of shared concepts can lead to hermeneutical injustice, the creation of shared concepts can be a way of correcting such an injustice. We can conceive of the confrontation of one group (in Lagewaard's example Dutch people who have never experienced racism) with the concepts of another group (in Lagewaard's example people living in the Netherlands who experience racism) as involving a conceptual disruption, though not one induced by technology. By challenging the first group's concept of racism, this confrontation can lead to uncertainty regarding the application conditions of this concept. This could initiate a process in which the concept is scrutinised and effort is undertaken to adjust the application conditions and develop related concepts that can capture the experiences of those suffering from racism.[16]

One way in which the injustice could be corrected is by one group (the group of those who have never experienced racism) adopting the concepts of the other. That might be challenging though since the lack of shared concepts is due to very different experiences. Presumably for it to happen members of both groups would have to engage in joint activities during which they get to know each other better and come to reach an understanding of each other's perspective. A possible way involves embodiment exercises in which people are asked to embody another person and interact with the other participants, thereby experiencing the situation and the relationships involved from their embodied perspective. This could also contribute to the development of new concepts that members of both groups can relate to, which is another way of correcting an injustice.

[15] "Testimonial injustice" refers to cases in which identity prejudices lead to the deflation of a testifier's credibility in the judgement of the person receiving the testimony (Fricker, 2007, pp. 20–28).

[16] For the history of the concept "racism" see Frederickson, 2002, referenced by Congdon, 2023.

7 Conclusion

In this paper, I connected the literature on deep disagreement to the literature on technology-induced deep disruption, exploring the complex relationship between three phenomena: certainty, deep disagreement, and deep disruption. I argued that all three phenomena are located at the same level– the level of bedrock practices– and that the fundamentality of the objects of certainty, deep disagreements and deep disruptions should be understood in terms of relationality and interconnectedness. Looking at two examples, I showed how moral deep disagreements can occur through deep technology-induced disruption and sketched how common ground could be re-established. Common ground consists not only of shared beliefs and preferences but also of shared concepts and agreement on how to apply them. I argued that in situations of conceptual disruption, a deep disagreement is not correctly described as a disagreement involving conflicting moral certainties. It is not a disagreement between people with fundamentally different world-pictures (unlike the case of abortion). Rather, such situations should be described as disruptions of the fundamental concepts that constitute a shared world-picture.

As I hope to have shown, the notion of technology-induced deep disruption can contribute to a richer understanding of the notion of moral deep disagreement. In turn, the concepts of certainty and deep disagreement can contribute to a richer conceptualisation of the phenomenon of deep technology-induced disruption. I argued that moral deep disagreement and disruption can lead to moral progress, by prompting a process of moral articulation, during which new moral concepts are developed that enable the experience and identification of hitherto elusive forms of injustice. Moreover, instances of hermeneutical injustice due to a lack of shared concepts could be resolved by creating conceptual common ground. Developing this line of reasoning further will be the subject of future research.

Acknowledgements I am grateful to Nigel Pleasants, Samuela Marchiori, and the members of the *Foundations & Synthesis Research Line* of the ESDiT programme for their valuable feedback on an earlier version of this paper. I would also like to thank the participants in the *Human-Technology-Relations Colloquium* of the University of Twente for their helpful comments and questions on the occasion of a presentation of the ideas presented in this paper. Finally, many thanks to two anonymous reviewers for their constructive feedback, which helped me improve the paper.

Funding This work is part of the research programme Ethics of Socially Disruptive Technologies, which is funded through the Gravitation Programme of the Dutch Ministry of Education, Culture, and Science, and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

Declarations

Competing interests The author has no competing interests to declare that are relevant to the content of this article.

References

Coliva, A., & Doulas, L. (2022). What philosophical disagreement and philosophical skepticism inge on. *Synthese*, *200*, 251. https://doi.org/10.1007/s11229-022-03735-6

Congdon, M. (2023). *Moral articulation: On the development of new moral concepts*. Oxford University Press.

Fogelin, R. J. (2005). The logic of deep disagreements. *Informal Logic*, *25*(1), 3–11.

Frederickson, G. M. (2002). *Racism: A short history*. Princeton University Press.

Fricker, M. (2013). Epistemic justice as a condition of political freedom? *Synthese*, *190*(7), 1317–1332. https://doi.org/10.1007/s11229-012-0227-3

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Goodman, R. B. (1982). Wittgenstein and ethics. *Metaphilosophy*, *13*, 138–148.

Hermann, J. (2015). *On moral certainty, justification, and practice: A wittgensteinian perspective*. Palgrave Macmillan.

Hopster, J., & Löhr, G. (2023). Conceptual engineering and philosophy of technology: Amelioration or adaptation? *Philosophy and Technology*. https://doi.org/10.1007/s13347-023-00670-3

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

Hopster, J. (2023). Conceptual disruption and the ethics of technology. In van de I. Poel, et al. (Eds.), *Ethics of socially disruptive technologies: An introduction* (pp. 141–162). Open Book.

Lagewaard, T. J. (2021). Epistemic injustice and deepened disagreement. *Philosophical Studies*, *178*, 1571–1592. https://doi.org/10.1007/s11098-020-01496-x

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

Löhr, G. (2022). Linguistic interventions and the ethics of conceptual disruption. *Ethical Theory and Moral Practice*, *25*, 835–849. https://doi.org/10.1007/s10677-022-10321-9

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

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

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

Metz, T. (2007). Towards an African moral theory. *The Journal of Political Philosophy*, *15*(3), 321–341. https://onlinelibrary.wiley.com/doi/https://doi.org/10.1111/j.1467-9760.2007.00280

Moore, G. E. (1959a). A defence of common sense. In Moore, G.E. (Ed.) *Philosophical Papers* (pp. 32–59). Allen and Unwin.

Moore, G. E. (1959b). Proof of an external world. In Moore, G.E. (Ed.) *Philosophical Papers* (pp. 127–50). Allen and Unwin.

Moyal-Sharrock, D. (2004). *Understanding Wittgenstein's* On Certainty. Palgrave Macmillan

Nickel, P., Kudina, O., & van de Poel, I. (2022). Moral uncertainty in echnomoral hange: Bridging the explanatory gap. *Perspectives in Science*, *30*(2), 260–283. https://doi.org/10.1162/posc_a_00414

Nickel, P. (2019). The Ethics of uncertainty for data subjects. In J. Krutzinna, & L. Floridi (Eds.), *The ethics of medical data donation* (Vol. 137, pp. 55–74). Springer. (Philosophical Studies Series.

Nyholm, S. R., & Frank, L. E. (2017). From sex robots to love robots: Is mutual love with a robot possible? In J. Danaher, & N. McArthur (Eds.), *Robot sex: Social and ethical implications*. The MIT Press. https://doi.org/10.7551/mitpress/10718.003.0019

Pleasants, N. (2008). Wittgenstein, ethics and basic moral certainty. *Inquiry 51*(3), 241–267. https://doi.org/10.1080/00201740802120673

Pritchard, D. (2021). Wittgensteinian hinge epistemology and deep disagreement. *Topoi*, *40*, 1117–1125. https://doi.org/10.1007/s11245-018-9612-y

Ranalli, C. (2021). What is deep disagreement? *Topoi*, *40*, 983–998. https://doi.org/10.1007/s11245-018-9600-2

Stroud, S. (2019). Conceptual disagreement. *American Philosophical Quarterly*, *56*(1), 15–28. https://doi.org/10.2307/45128640

Van de Poel, I., Frank, L. E., Hermann, J., Hopster, J., Lenzi, D., Nyholm, S., Taebi, B., & Ziliotti, E. (Eds.). (2023). *Ethics of socially disruptive technologies: An introduction*. Open Book.

Williams, M. (1999). *Wittgenstein, mind and meaning*. Routledge.

Wittgenstein, L. (1968). *Philosophical investigations*. Translated by G.E.M. Anscombe. Third ed. Basil Blackwell.

Wittgenstein, L. (1972). *On certainty*. Translated by D. Paul and G.E.M. Anscombe. Edited by G.E.M. Anscombe & G.H. von Wright. Harper & Row.

Wittgenstein, L. (1978). *Remarks on the foundations of mathematics. Translated by G.E.M. Anscombe. Edited by G.H. von Wright e.al. Third, revised and reset ed. Basil Blackwell.*

Wittgenstein, L. (1993). Cause and effect: Intuitive awareness. In Klagge, J.C. & Nordmann, A. (eds.) *Philosophical occasions 1912–1951* (pp. 370–426). Hackett Publishing Company.

Authors and Affiliations

Julia Hermann[1]

✉ Julia Hermann
j.s.hermann@utwente.nl

[1] Department of Technology, Politics, and Society; Philosophy Section, University of Twente, Enschede, The Netherlands