Morality in the age of artificially intelligent algorithms

Moser, Christine; den Hond, Frank; Lindebaum, Dirk

published in
Academy of Management Learning and Education
2022

DOI (link to publisher)
10.5465/amle.2020.0287

document version
Publisher's PDF, also known as Version of record

document license
Article 25fa Dutch Copyright Act

© *Academy of Management Learning & Education*
2022, Vol. 21, No. 1, 139–155.
https://doi.org/10.5465/amle.2020.0287

MORALITY IN THE AGE OF ARTIFICIALLY INTELLIGENT ALGORITHMS

CHRISTINE MOSER
Vrije Universiteit Amsterdam

FRANK DEN HOND
Hanken School of Economics
Vrije Universiteit Amsterdam

DIRK LINDEBAUM
Grenoble Ecole de Management

This article starts from the premise that human judgment is intrinsically linked with learning and adaptation in complex sociotechnological environments. Under the illusory veneer of retaining control over algorithmic reckoning, we are concerned that algorithmic reckoning may substitute human judgment in decision-making and thereby change morality in fundamental, perhaps irreversible ways. We present an ontological critique of artificially intelligent algorithms to show what is going on "under their hood," especially in cases when human morality is already co-constituted with algorithmic reckoning. We advance a twofold call for (in)action. First, we offer a call for *in*action as far as the substitution of judgment for reckoning through our teaching in business schools and beyond is concerned. Second, we advance a reinvigorated call for action—in particular, to teach *more* pragmatist judgment in our curricula across subjects to foster social life (rather than stifle it through algorithmic reckoning).

Numerous studies converge on the notion that learning can be defined "as the basic process of human adaptation" (Kolb & Kolb, 2009: 42). Learning is often understood beyond narrow cognitivism to include an integration of a person's thinking, feeling, and behaving (Greene & Haidt, 2002; Kolb & Kolb, 2005). It thus involves the "continuing reconstruction of experience" (Dewey, 1897: 13) to adapt to instances of conflict, ambiguity, disagreement, and difference that we encounter as members of society. However, *when*, *how*, and *why* we adapt is a matter of judgment, because it depends on the contingencies of time, space, and social context. We define "judgment" as making decisions that take into account the social and historical context and different possible outcomes, with the aim "to carry an incomplete situation to its fulfilment" (Dewey, 1916: 362). Judgment implies not only reasoning but also, and importantly so, capacities such as imagination, reflection, examination, valuation, and empathy. Therefore, it has an intrinsic moral dimension (cf. Shotter & Tsoukas, 2014). As such, learning is associated with judgment at two instances: (1) while forming judgment and (2) when reflecting on the outcome of acting upon judgment.

Emphasizing the link between learning and judgment matters considerably vis-à-vis the rapid proliferation of artificially ⌈intelligent⌉[1] algorithms in

Our sincere gratitude goes to former editor in chief Bill Foster for an excellent editorial steer, and to the two reviewers for providing a thoughtful and stimulating set of reviews. We are also grateful for the constructive and insightful comments from participants of the 15th Organization Studies Summer Workshop, and the 1st Organization Theory Winter Workshop. In addition, we would like to thank Dirk Deichmann, Mariel Jurriens, and Laura Schons for their feedback and invaluable support in writing this manuscript. All remaining issues are ours. Note that this article was submitted before Dirk Lindebaum became an associate editor for this journal. The first two authors contributed equally to the manuscript and share first authorship. The practical implications of our theorizing are elaborated on in a forthcoming article in the MIT Sloan Management Review issue in Spring 2022.

[1] We follow a convention proposed by Smith (2019: 50): "I will mark with corner quotes ('⌈' and '⌉') terms we standardly apply to computers that I believe rely on our interpretation of the semantics of the action or structures, rather than anything that the system itself can be credited

management (e.g., Kellogg, Valentine, & Christin, 2020; Newlands, 2021; Raisch & Krakowski, 2021). In this article, we offer the strong thesis that we are at risk, *now*, that these algorithms change, perhaps irreversibly so, our morality in fundamental ways by suppressing judgment in decision-making. We develop and use such algorithms to facilitate and enhance decision-making, harboring the illusion that we, human beings, are in control and can develop algorithms to emulate judgment by being aligned with, reflecting, and espousing our morality.[2] However, as we increasingly rely on artificially ⌈intelligent⌉ algorithms in decision-making, we risk mistaking "reckoning"—"the calculative rationality of which present-day computers … are capable" (Smith, 2019: 110) by processing data through an accumulation of calculus, computation, and rule-driven rationality—for judgment. Mistaking reckoning for judgment, presuming them to be ontologically similar whereas they are not, may impoverish our morality because, if we come to believe reckoning and judgment to be the same, the former might eventually replace the latter (cf. Lindebaum, Vesa, & den Hond, 2020). This process is already ongoing, as we argue in this article. In light of this process, we are, in fact, far less in control of algorithms than currently recognized. Therefore, and against current trends enthusiastically professing otherwise, we set out to shine a light on the unexamined processes through which we risk losing control over artificially ⌈intelligent⌉ technology. If we do lose control, we risk fashioning ourselves and our social life in the image that the technology is creating of us. In management and beyond, we need judgment, not reckoning—especially if we want to remain adaptive in the sociotechnological world that we now inhabit.

Before we proceed, some definitions and delimitations are in order. First, "algorithms" are "precise recipes that specify the exact sequence of steps required to solve a problem" (MacCormick, 2012: 3). Computers run on algorithms. Our focus is on those algorithms that make computer systems "artificially ⌈intelligent⌉." Artificial ⌈intelligence⌉ (AI) is shorthand language for a set of complex algorithms that have been under development since the early 1950s, initially to simulate human intelligence, and later to support or even take over and autonomously execute tasks in a complex environment. Current AI is characterized by the ability to improve its own performance through techniques known as "machine ⌈learning⌉" (e.g., Sun, 2014; Mitchell, 2019) and, in many ways, such AI systems have by now become "fundamental features of contemporary organizing" (Glaser, Pollock, & D'Adderio, 2020: 3). More generally, current AI systems are able to operate autonomously, to adapt—that is, to ⌈learn⌉—in response to environmental stimuli and feedback, and to interact with the external world through exchange of information with human and other non-human agents (e.g., Alonso, 2014; Dignum, 2019). Second, we embrace an ontological vantage point (Lawson, 2019), because it enables us to examine the ontological assumptions underlying algorithmic reckoning and human moral judgment. We introduce Flusser's (2000) notion of the "technical image" to the management learning community as a way to explain how the outputs of algorithms are abstractions that distort our understanding of the phenomena from which they abstract. Finally, we interpret the process through which judgment is assimilated into algorithmic reckoning as "ontological assimilation."

Our argument is of vital importance for the management learning community for two reasons. First, the way that AI is currently being taught in business schools highlights commercial opportunity (e.g., in an entrepreneurial discourse of start-up culture, and without much attention to wider societal implications thereof). By contrast, Vesa and Tienari (2020) argued how AI discursively allows its owners to wield power and exert control in direct and indirect ways over citizens, customers, and societies, which is justified by an ideology of rationality but otherwise escapes accountability. In this way, they argued that "AI functions as an ideology as it manufactures normative idea(l)s of social reality into self-evident truths, benefitting

with understanding or owning. For instance, image or face ⌈recognition⌉, algorithmic ⌈decision-making⌉, and so on." Smith's proposal is, accidentally, in full accordance with the Academy of Management's "Style Guide for Authors" in its banishing of anthropomorphisms ("Do not describe inanimate entities (models, theories, firms, and so forth) as acting in ways only humans can act"; Academy of Management, 2021: 2).

[2] In this regard, a recent white paper published by the European Commission states: "Artificial intelligence is developing fast. It will change our lives by improving health care (e.g., making diagnosis more precise, enabling better prevention of diseases), increasing the efficiency of farming, contributing to climate change mitigation and adaptation, improving the efficiency of production systems through predictive maintenance, increasing the security of Europeans, and in many other ways that we can only begin to imagine" (European Commission, 2020: 1).

some [e.g., the future leaders of industries] at the expense of others" (Vesa & Tienari, 2020: 10). Such critique is needed, but not enough, as it leaves the AI itself "black boxed." Our article helps to demystify AI: laying bare the underlying mechanisms of AI decision-making will help teachers and students to better understand what is going on "under the hood" of AI. Second, we provide teachers with a perspective and heuristics to express criticism. While much critique of AI is inspired by Hollywood science fiction dystopias (Broussard, 2018), our unpacking of the role of AI in decision-making makes concrete the claim that current AI already presents a risk for society. Our argument, therefore, has considerable implications for how, and on the basis of what contents, we teach a range of courses in business schools, such as business ethics, decision-making, or individual and organizational learning (Balasubramanian, Ye, & Xu, 2020; Hibbert & Cunliffe, 2015; Loon, 2020).

Our claim that increased reliance on reckoning—and, ultimately, the substitution of reckoning for judgment—may result in an impoverished human morality requires attention to central issues, which helps structure our article. First, decision-making and morality have a recursive relationship; morality influences decision-making as much as decision-making influences morality. Second, both judgment and reckoning may contribute to decision-making and thereby constitute morality. Third, having thus related judgment and reckoning to morality, we explore the divergent ontological assumptions underlying judgment and reckoning in decision-making. From these starting points, we sketch and discuss three scenarios of how judgment and algorithmic reckoning may play out in decision-making, and, through that, affect morality. The first scenario is a dystopian extension of the current trend to rely on algorithmic reckoning in decision-making, which we associate with a process of "ontological assimilation." The second scenario examines the current discourse of "responsible AI" to argue that the promise of this discourse is exaggerated in light of the possibility of ontological assimilation. The third scenario starts from the acknowledgment that algorithmic reckoning is already affecting morality through its material agency (Introna, 2014). We invoke Flusser's (2000) technical image to explain how this works. This, in turn, motivates our twofold call. One is a call for *in*action; literally, a call to *in*action as far as the substitution of judgment for reckoning is concerned. The other is a call for action, a reinvigorated call to teach *more* pragmatist judgment in our curricula across subjects.

DECISION-MAKING AND MORALITY: A RECURSIVE RELATIONSHIP

Decision-making is an elusive and ambiguous concept that resists unequivocal definition. At its core, "decision-making" is about developing and selecting a course of action out of a number of alternatives. It is thus related to choice (Brunsson & Brunsson, 2017) and deliberation (Habermas, 1993, 1996), but also to the upholding of choice (Bachrach & Baratz, 1963). Although decisions can thus be made consciously and explicitly, they may also accidentally or unreflectively "happen" (Cohen, March, & Olsen, 1972; March, 1994). However they are viewed, decisions are often only accounted for or justified after they have been "made" (e.g., Haidt, 2001). Computer systems are said to make decisions when the output of their reckoning, or calculus, is taken as such (cf. "calcucision"; Lindebaum et al., 2020). However, there is a normative understanding that decisions are "better" when they are, or can be, justified and accounted for on the basis of some appropriate substantive value orientation (Weber, 1968); that is, decision-making and morality are related.

"Morality" encompasses socially developed norms and practices for regulating conflicting interests. As such, it is circumscribed in space and time and has a general function or role in regulating social life (Dewey & Tufts, 1932; Lindebaum, Geddes, & Gabriel, 2017; Wong, 2006). As a normative concept, the prevailing morality informs decision-making by suggesting appropriate ways to act in an environment; it suggests, informs, or prescribes which norms to adhere to, and how. Yet, the authority that is attributed to morality—in terms of the substance of its norms, how to adhere to them, and the level of stringency of its demands—varies across space and over time. Although morality is experienced as a normative concept (and thus seems to be stable), it evolves over time. This is because it emerges from retaining satisfactory ways of dealing with conflicting interests that stem from novel experiences and conditions, or from new sociotechnical possibilities, for example. In this way, decision-making—the selection of a particular course of action—may lead to the adoption of new ways of regulating social life and thereby affect morality. Morality, thus, not only informs decision-making but is also affected by decision-making: decision-making and morality have a recursive relationship.

One does not have to delve deeply into the recent COVID-19 crisis to see not only how morality and decision-making have a recursive relationship, but also how both reckoning and judgment played their roles in figuring out "what is best and wise to do" (Dewey, 1922: 190) in the face of the pandemic. Decision-making relies on the processing and evaluation of information ("data") relevant to an ambiguous, troubled, problematic, or puzzling situation. Reckoning and judgment both feed into decision-making, but in quite different ways. In the case of COVID-19, judgment had the upper hand in some countries, in the sense that there was both a continuing scrutiny of the relevance and validity of the data that were fed into models that attempted to predict the development of the pandemic (e.g., Schumann, 2020), and a continuing weighing of the social and economic consequences of measures to contain the spread of the virus. In other countries, reckoning had the upper hand in the sense that facts, such as changes in "R" and the capacity of the health care system, dictated which measures were taken. We conclude this section by reiterating that decision-making—whether based on human judgment, algorithmic reckoning, or a combination thereof—not only expresses but also constitutes morality.

JUDGMENT AND RECKONING IN DECISION-MAKING: DIVERGENT ONTOLOGICAL ASSUMPTIONS

We understand "ontology"—literally, the study or the knowledge of the nature of being—as a set of assumptions that inform answers to fundamental questions about the nature and reality of phenomena. Understanding the nature of the phenomena of interest is essential to virtually everything we do as social scientists (Watson, 2013). It is, as Lawson (2019: xi) put it, "inescapable; we all make assessments of the nature and constitution of social reality continuously already just in order to get by." And yet, we often do so rather implicitly, and without full consciousness, such that ontological assumptions remain hidden from view (Lawson, 2019). Making these ontological assumptions explicit helps us to point out the different consequences that flow from judgment and reckoning in relation to decision-making. Specifically, we associate judgment with a pragmatist ontology and reckoning with a principled ontology.

Judgment is about "finding out what the various lines of possible action are really like … to see what our resultant action would be like if it were entered upon" (Dewey, 1922: 190), such that an informed decision can be made in a given situation with an eye to improving that situation. It is about what is appropriate, right, good, fair, or just to do in an ambiguous, troubled, problematic, or puzzling situation, having explored and considered the various characteristics of that situation and having (creatively) developed and (carefully) evaluated multiple options with regard to their respective potential to "better" that situation. Judgment, therefore, requires imagination, reflection, empathy, and valuation. In judgment, it is acknowledged that data are value laden, and that the identification of which values are relevant for decision-making is an inherent part of the process (cf. Dewey, 1939). Moral considerations thus inescapably come into play when developing judgment because they cannot be excluded or separated from the very situation that demands judgment (cf. Dewey, 1922; Dewey & Tufts, 1932). Owing to its ambiguous, evolving, and plural nature (Dewey, 1922), a pragmatist ontology assumes that the world can never be fully understood and predicted: it demands ontological experientialism and epistemological fallibilism (Martela, 2015; cf. Simpson & den Hond, 2021). Consequently, understanding is therefore always "perspectival" in the dual meaning of "originating from a perspective" and "being oriented toward a perspective."

By contrast, reckoning is the processing of data through calculus and formal rationality (Lindebaum et al., 2020). It relies on data as correct representations of reality ("facts"), and values can only find their place in reckoning as stable ex ante givens, indeed a form of "data." Driven by predefined rules and goals, reckoning is insensitive to context and time. Accordingly, reckoning can only proceed from a view that sees the world as principled and discrete—hence our labeling of the ontology underlying reckoning as "principled." In this view, the world is understood in terms of logical and "objective" relationships that are fully and unambiguously defined. Its matching epistemology is premised on "an approach to knowledge that seeks to deduce knowledge from first principles or a priori general ideas—principles acquired or obtained prior to any actual human experience" (Azelvandre, 2001: 170). Data and information are seen as unproblematic representations of the world, rather than—from a pragmatist viewpoint—as discriminatively selected, assembled, and created with the purpose of "affording signs or evidence to define and locate a problem, and thus give a clew [*sic*] to its resolution" (Dewey, 1929: 178). In sum, the ontological assumptions underlying judgment and reckoning thus sharply diverge.

THREE SCENARIOS OF THE INTERACTION BETWEEN HUMAN JUDGMENT AND ALGORITHMIC RECKONING

Before we proceed, we need to unpack how and why current AI is associated with a principled ontology. There are currently two forms of machine ⌈learning⌉: (1) supervised and (2) reinforcement. Mitchell (2019) likens supervised ⌈learning⌉ to a kind of behaviorist training in which humans reward or punish the AI system to train it. Workers on platforms such as Mechanical Turk feed the AI system with a large number of examples, labeled as correct or incorrect in relation to a criterion ("Is there—'yes' or 'no'—a cat in the picture?"): this is the "training set." On this basis, the system is able to ⌈learn⌉ how to classify previously not analyzed input with a certain amount of probability. Applications of AI systems based on supervised ⌈learning⌉ are found in image ⌈recognition⌉ and chatbots. Reinforcement ⌈learning⌉ can be explained as a kind of operant conditioning (Mitchell, 2019) in optimizing action in response to feedback. The AI system takes a series of (random) actions in a defined environment, each of which provokes an immediate response from that environment. The system uses these responses as data in calculating the relative progress of each action toward a predefined goal and thus, by retaining the "best" action, is able to determine which next action brings it closer to the predefined goal. Reinforcement ⌈learning⌉ applies a repeated trial and error model of ⌈learning⌉ to accomplish a goal set by humans, such as "beat Atari'[3] or "win a game of Go" (Silver et al., 2017). Applications of AI systems based on reinforcement ⌈learning⌉ are found in algorithms that offer consumers next-choice or next-purchase suggestions, such as in online stores, search engines, and streaming services.

These abilities—to ⌈identify⌉ patterns in huge amounts of unstructured data, in the case of supervised ⌈learning⌉, and to select a next move with a relatively high likelihood of contributing to a pre-established goal, in the case of reinforcement ⌈learning⌉—are both executed through repeated mathematical operations on digitized data; the former through optimization in pattern recognition, the latter through maximization. This reckoning depends on pre-coded information and predefined rules. Regardless of the amount or complexity of data or the type of machine learning applied, the AI cannot function without these predefined rules or pre-coded information. And, for these, human intervention is needed to define the purpose, the goal, and the categorization of the data (Mitchell, 2019; Smith, 2019). Underlying this logic is an understanding and utilization of data as if they were unequivocally correct representations of facts. In other words, data are being treated as if they were logical, discrete, and unambiguously defined—which is the realm of reckoning and principled ontology.

At this juncture, we can sketch three possible scenarios of the interaction between judgment and reckoning in decision-making and how each scenario relates to morality (Figures 1a–1c). Scenario A (Figure 1a) depicts a dystopian extension of the current trend in which algorithmic reckoning is already replacing judgment in decision-making; in this scenario, morality will eventually become algorithmic. The dystopia of this scenario has been elaborated in, for example, Lindebaum et al. (2020). Next, scenario B (Figure 1b) covers the current discourse around AI and ethics that seeks to control algorithmic reckoning in decision-making by informing, guiding, and steering the development and use of AI in line with human morality, under the presumption that AI can be controlled by judgment. We find this scenario naïve and myopic in its idealization of human control in the upholding of morality and neglect of ontological assimilation. Finally, scenario C (Figure 1c) represents what we believe is the actual condition in which decision-making is already co-constituted by judgment and reckoning. We argue that, paradoxically, the continuing pursuit of scenario B might move us toward scenario A, and that, to prevent this from happening, we need to acknowledge and take seriously scenario C: AI-informed decision-making is already changing our morality, due to the material agency of AI.

Scenario A: Algorithmic Reckoning in the Lead

> "Sir, my need is sore! Spirits that I've cited my commands ignore."
>
> —Johann Wolfgang von Goethe, "The Sorcerer's Apprentice"

[3] "Google DeepMind's Deep Q-learning Playing Atari Breakout" (Two Minute Papers, 2015, cited in Tegmark, 2017).

FIGURE 1
Three Scenarios of the Interaction between Human Judgment and Algorithmic Reckoning

Figure la: Algorithmic morality; reckoning substituting for judgment in decision-making

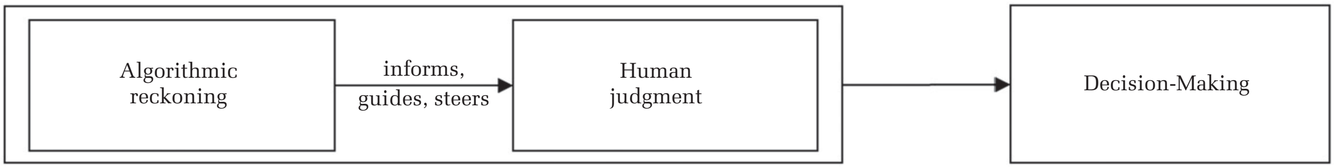


Figure 1b: Human morality; reckoning subservient to judgment in decision-making

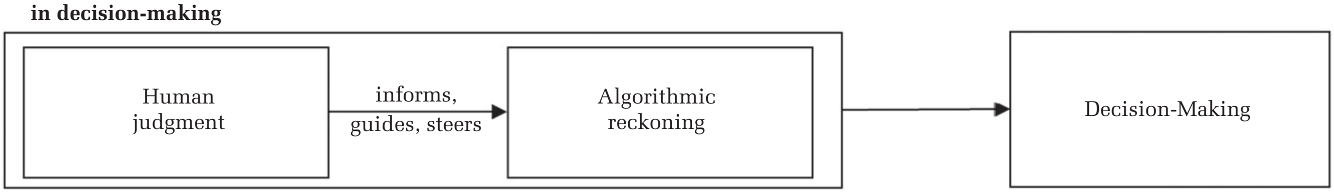
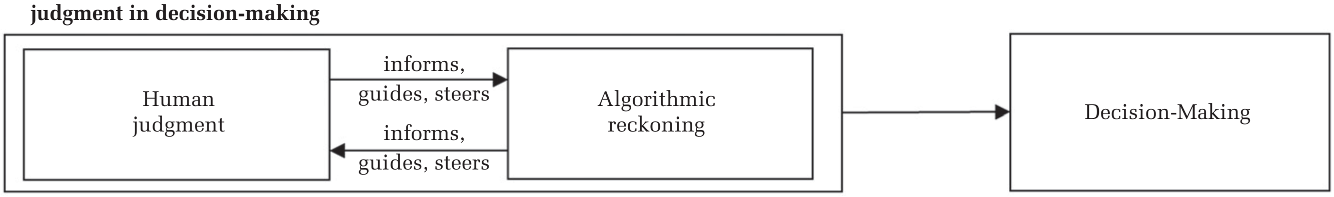


Figure lc: Co-constituted morality; a blending of reckoning and judgment in decision-making



In scenario A (Figure 1a), algorithmic reckoning informs, guides, and steers decision-making. Decision-making has effectively been relegated to AI systems. Imagined futures in which artificial ⌈intelligence⌉ has completely taken over the world may never materialize. However, much decision-making has already been left to the reckoning of AI systems, perhaps in the belief that it does not matter whether decisions are made through judgment or reckoning, or perhaps in the belief that the AI system has superior qualities over human beings in terms of speed and accuracy in processing data for decision-making. Although both judgment and reckoning may contribute to decision-making, there is an asymmetry in their roles: whereas judgment may include reckoning, reckoning in and of itself excludes judgment. It is because of this asymmetry that reckoning may eventually replace judgment in decision-making.

We refer to this process as "ontological assimilation," whereby judgment is being straightjacketed, curtailed, and amputated to produce reckoning. Elsewhere, this has been discussed as the transformation and subordination of substantive rationality to formal rationality (Lindebaum et al., 2020). Ontological assimilation is problematic because it entails a separation of fact and value, a separation that has been argued to be both impossible and undesirable (e.g., Dewey, 1939; Putnam, 2002). Rejecting this separation implies that moral issues can never be dealt with in an abstract, a priori manner. Instead, judgment requires that moral issues are treated as empirical questions of which values are at stake and how they may play out in improving that particular situation (Dewey, 1939). What this suggests, however, is that, for AI to be ⌈responsible⌉ or ⌈ethical⌉, judgment has to be molded in the mathematical language of algorithms—it has to become reckoning. Judgment must be assimilated to the principled ontology of reckoning—that is, be "ontologically assimilated."

Because of their reckoning, algorithms require a conceptualization of morality as independent of time and locale, as an objective phenomenon with universal validity, in accordance with a principled ontology. Therefore, we would have to define—in said universal fashion—the fundamental grounds of human morality prior to decision-making and "objectively" vis-à-vis the decision situation at hand

and, moreover, in a manner that algorithms can process it. Such a reconceptualization of morality would only be acceptable if there were agreement not only on morality being principled in the first place, but also on which of its multiple formulations is the correct one at this point in time. In addition, such agreement would have to be translated into the formal language of computer code. These conditions are in contradiction to the pragmatist ontological assumptions underlying judgment as described above. As Dewey (1916: 374) put it:

> The standard of valuation is formed in the process of practical judgment or valuation. It is not something taken from outside and applied within it—such application means there is no judgment.

Therefore, and by necessity, any prospect of "merging" judgment and reckoning, or of "bringing them into alignment," implies ontological assimilation.

Scenario B: Human Control over Decision-Making

There is much evidence of the detrimental consequences of our current reliance on algorithmic reckoning in decision-making (e.g., O'Neill, 2016; Redden, Brand, & Terzieva, 2020; Zuboff, 1988, 2019). For example, the unjustified and immoral consequences of its reliance on reckoning in decision-making—in the context of a program to counter fraud in a child benefit tax relief scheme—has most recently prompted the resignation of the Dutch government (Henley, 2021). Scenario B describes the efforts currently being made to counter or prevent from happening these undesired consequences of scenario A. Despite the mixed appraisal of benefits and risks associated with AI systems (e.g., European Commission, 2020), it is believed, in scenario B, that it is possible to "tame" the algorithm. Scenario B, therefore, describes the aim to restore the situation in which judgment informs, guides, and steers decision-making, with algorithms merely used as tools to aid decision-making. It starts from the premise that the risks associated with AI can be controlled by developing and designing it to serve human needs and values, and to support human morality.

And indeed, considerable effort is being made to develop "responsible AI"—applications of AI that are subservient to human needs and values—in response to widespread discussions in mass media, public policy, and academic circles about AI and ethics. For example, Tegmark (2017) emphatically argued for the need to develop "beneficial AI" in light of its unstoppable and inevitable further development. Ames (2018: 3) pointed to research that discusses AI as an artifact of culture, such that values and interests are by necessity being embedded in AI systems through their programming, training, and use: "algorithms have everything to do with the people who define and deploy them, and the institutions and power relations in which they are embedded." This would enable the possibility of controlling AI through these "people and institutions," for example through the formulation of codes of conduct for the development and use of AI systems.[4] Such views of responsible AI treat AI as a tool, on par with other tools that are essentially extensions of the human body, such as lenses (to extend the view of the eyes), thermometers (to extend the sense of the skin), and spoons and screwdrivers (to extend the dexterity of the hands).

But is the premise underlying scenario B in any way realistic or viable? We are skeptical. In this section, we seek to critically examine scenario B in light of our thesis of ontological assimilation. We do so through a discussion of Dignum's (2019) state of the art overview of how AI can be made "responsible."[5] Her perspective on responsible AI hinges on three interrelated approaches to designing and using AI, which she labels "ethics *in* design," "ethics *by* design," and "ethics *for* design(ers)."

---

[4] By 2019, both public and private organizations—most of them originating from Europe and North America—had already published well over 80 non-legally binding formulations of principles and guidelines for AI ethics, predominantly focusing on the moral obligation to prevent harm (Jobin, Ienca, & Vayena, 2019). In a similar kind of analysis, Floridi and Cowls (2019) identified five principles of AI ethics: "beneficence," "non-maleficence," (human) "autonomy," "justice," and "explicability." Whereas the first four are similar to principles of bioethics, "explicability" is specific for AI. "Explicability" has a dual meaning; it is to be understood as an answer to the question "How does it work?" and as an answer to the question "Who is responsible for the way it works?" (Floridi and Cowls, 2019: 8). The principle of explicability is deemed significant for ensuring and enhancing trust in AI systems (Glikson & Woolley, 2020).

[5] Dignum's (2019) understanding of "responsible AI" is echoed by many others. Further useful sources include Anderson and Anderson (2011), Martin (2019), and Pereiro and Lopes (2020), as well as several manuscripts by Floridi and colleagues (e.g., Floridi, 2019; Floridi et al., 2018).

In the ethics-*in*-design approach, ethical principles and moral values are translated into design requirements for AI systems. Beyond adhering to general principles of AI ethics,[6] the approach centers on three steps (Dignum, 2019: 62):

> (1) the identification of societal values, (2) deciding on a moral deliberation approach (e.g., through algorithms, user control, or regulation), and (3) linking values to formal system requirements and concrete functionalities.

On the upside, this approach makes visible what is inevitable—that AI systems do espouse values and that the intentionality of AI systems is intimately connected to, and derived from, human intentionality and agency through the act of design (Johnson, 2006).[7] Other than that, this approach can only proceed through ontological assimilation.

If ethics-*in*-design is a "top-down" approach, then ethics-*by*-design is a "bottom-up" approach (Etzioni & Etzioni, 2017); it refers to designing AI systems in such a way that they *themselves* acquire the capacity of moral ⌈reasoning⌉ in producing their output. Current AI systems approach this capacity in either of two ways: (1) algorithmically or (2) in a random manner. In the latter approach, the AI system randomly ⌈chooses⌉ among a set of preprogrammed options. The justification is in the claim that, "if it is ethically problematic to choose between two wrongs, a possible solution is to simply *not* make a deliberate choice" (Dignum, 2019: 87, emphasis added). Judgment is replaced by a Monte Carlo function. The former, algorithmic approach aims to fully incorporate moral reasoning into the system through the autonomous evaluation of the moral and societal consequences of its decisions (Wallach & Allen, 2009), such as would be needed in autonomous vehicles when facing situations of unavoidable harm. In practice, this means either the formalization of some combination of principled ethical theories, or using empirically measured social preferences vis-à-vis a morally aporetic situation as a proxy for judgment, such as in the MIT Moral Machine experiment (Awad et al., 2018). It may be envisioned that these preferences can function as a training set for supervised ⌈learning⌉ (and may even accommodate variations in the average espoused preferences in different parts of the world). However, this amounts to the adoption of the naturalistic fallacy—the immediate and indiscriminate transition of "is" to "ought"—and a reduction of judgment to a popular vote.

Finally, the ethics-*for*-design(ers) approach is about the "mechanisms that can ensure that all [humans] involved will indeed take the responsible route" (Dignum, 2019: 93). It includes the introduction of governance structures and codes of conduct to guide the professionals developing and using AI systems. Without either of the two other approaches, this approach puts the onus of responsible AI on AI professionals instead of imposing demands on AI. Yet, there is ample evidence on the inefficacy of codes of conduct and other governance systems in controlling human behavior (see, e.g., in the domain of organizational ethics, Helin, Jensen, Sandstrom, & Clegg, 2011; Nyberg, 2008).

We conclude that the third approach bypasses the issues at stake, whereas, for various reasons, the first and second approaches fall short of the premise of "responsible" AI, if responsibility is associated with judgment. Other commentators have come to similar conclusions. In a critical analysis that starts from "the inner structures of ethical philosophies used by humans," Etzioni and Etzioni (2017: 408, 404) concluded that "there is little need to teach machines ethics even if this could be done in the first place." Likewise, Bryson (2018) concluded that, even if it were technically possible to create AI systems that would meet contemporary requirements for moral agency, it is neither necessary nor desirable that we should do so. For us, as well as for Etzioni and Etzioni (2017) and Bryson (2018), the point is to understand at a fundamental, ontological level the agencies and limitations of AI.

Scenario C: Co-Constituted Decision-Making and Morality

We believe that a more realistic view is depicted in Figure 1c. At the core of this scenario C is the recognition that decision-making, and thereby also morality, is *co-constituted* with judgment and reckoning. Our previous arguments about the recursive relationship between decision-making and morality and about ontological assimilation point to the argument that "technology can do things with or through humans as such" (Introna, 2014: 34). Concepts such as "non-human actants" (Latour, 2005) and the "affordances" of things (Gibson, 1979) have been

[6] See Jobin et al. (2019) and Floridi and Cowls (2019); Dignum lists accountability, responsibility, and transparency.

[7] The argument is foundational of the social construction of technology tradition in technology studies (Bijker, Hughes, & Pinch, 1987).

developed to explain how technology can be "agentic" (Murray, Rhymer, & Sirmonet, 2021). Extending such explanations, and in line with the sociomaterial tradition (e.g., Leonardi, 2012; Moser, Reinecke, den Hond, Svejenova, & Croidieu, 2021; Orlikowski, 2007), we advocate the claim that AI affects decision-making in such a way that morality is *co-constituted* with algorithms (Introna, 2014). Underlying this claim is the view that the social and the technical are inseparable because "agency is not an attribute of the human or the technical as such but rather the outcome of intra-action" (Introna, 2014: 5). In a very general sense, technology changes our outlook onto the world. It changes our sense of possibility, and, as the newly possible becomes routine, it also changes our sense of "what ought to be."

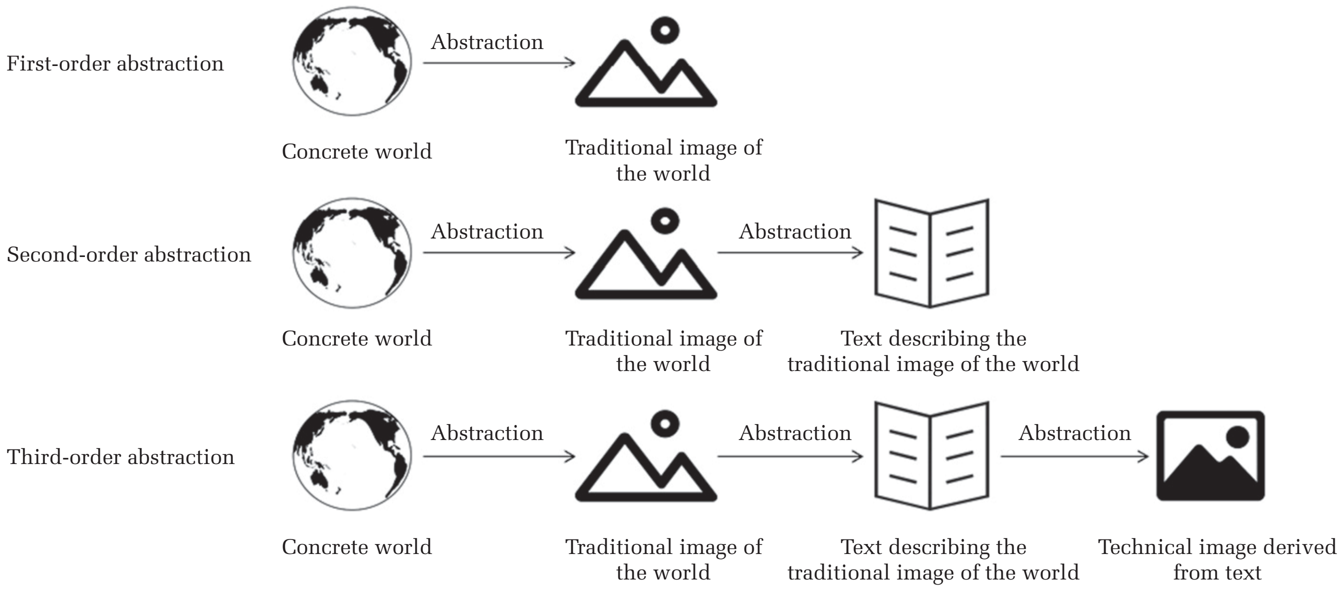
To illustrate how AI already changes our morality, we provide two (out of many possible) examples. One example is from the domain of health, the other one from digital assistants. To start with, people increasingly use "wearables" (small devices, like a watch, that collect data and communicate with smartphones and tablets) to monitor their "health"—which has now become a summary of the various data points that the AI embedded in the smartphone app returns. A numerical definition of "health" is embedded in the algorithm on the basis of which it entices us to improve our bodily condition to emulate a predefined standard (cf. Elmholdt, Elmholdt, & Haahr, 2021, for a work-related example). Instead of assessing health in a way that does justice to the individual body and well-being, wearables reduce us to what can be captured as quantifiable data. It denies a moral understanding of health as a phenomenological experience (Gadamer, 1996) and seeks a disciplining of the body to external standards (Foucault, 2008). Our second example stems from Bonfert, Spliethöver, Arzaroli, Lange, Hanci, and Porzel (2018), who described how digital AI assistants such as Alexa and Siri can become role models. Instead of saying "please," children learned to use a "command voice" that was perceived as rude by their parents but required by the digital assistant. In examples such as these, AI affects morality, as it changes how we regulate our social life. We already find ourselves in situations in which we nurse the illusion of being in control of judgment. We *think* that we control how we use wearables whereas studies show how people change their very outlook on life because of the technology (Balconi, Fronda, Venturella, & Crivelli, 2017); we *think* that we control recommender algorithms on Amazon, whereas studies show that people are so easily influenced by online reviews (Zhao, Stylianou, & Zheng, 2018), regardless of their authenticity or "morality."

Scenario C may digress into scenario A, if we leave things as they are. However, it does not necessarily *have* to do so, if we understand how exactly AI operates "under the hood." Such understanding does not have to be able to retrace how, say, an AI system for image [recognition] categorizes every single picture as to whether or not it depicts a cat—as is the idea of "explainable" AI—but it does need to have a detailed understanding of what happens in the process of the digitization of data—the transformation of qualia into quanta, their subsequent processing, and the production of output. If we do have such an understanding, we have a choice of when, why, and to what extent we can enroll algorithmic reckoning into our decision-making. In the following, we draw on Flusser's (2000, 2011) idea of the "technical image" to offer such understanding. Telling in this regard is his observation that:

> The encoding of technical images … is what is going on in the interior of this black box and consequently any criticism of technical images must be aimed at an elucidation of its inner workings. As long as there is no way of engaging in such criticism of technical images, we shall remain illiterate. (Flusser, 2000: 16)

Flusser positions himself in the tradition of Heidegger's (1977) critique of technology, where the idea of the "apparatus" has been developed. According to Agamben (2009: 14), an "apparatus" is "literally anything that has in some way the capacity to capture, orient, determine, intercept, model, control, or secure the gestures, behaviors, opinions, or discourses of living beings." Flusser, in this same tradition, examined that which Agamben backgrounded: modern science in its multiple ways of producing "technical images" through apparatuses. In the specific context of AI, Flusser's discussion of this latest offspring of the "second major revolution in the history of humankind: the invention of the technical image" (Flusser, 2000, 2011), is, therefore, of critical relevance. For our purposes, it is important to note that we use the terms "apparatus" and "technical image" as a simile: "apparatus" refers to an AI system that runs on algorithms, and "technical image" to what the system produces, subsequent to its capturing and processing of digitized data. The reason for us to invoke the simile of the technical image is that

FIGURE 2
Orders of Abstractions



it lets us realize how AI and humans are already entangled in their co-constitution.[8]

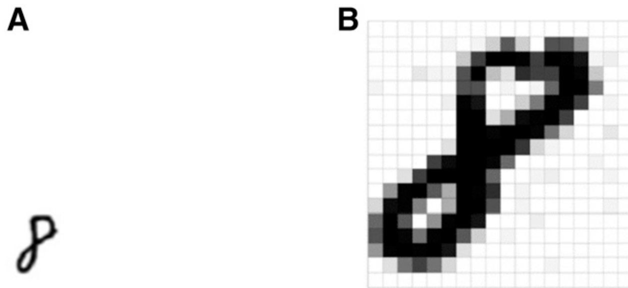
Flusser (2000, 2011) made a distinction between traditional images and technical images. A traditional image—such as a drawing, a painting, a sculpture—is a depiction of some object(s) or idea(s) in some way based on the experience of its human creator. It is a first-order abstraction as it expresses a meaningful, or semantic, relationship to the object(s) or idea(s) it depicts. A technical image, by contrast, does not have such a semantic relationship; instead, it is a visualization of a *computed* transformation of digitalized data of some object or idea (which itself was made possible through a second-order abstraction and theorizing of traditional images in the form of scientific text). Thus, technical images are abstractions of the third order. Whereas traditional images are "meaningful surfaces," technical images are "mosaics assembled from particles" (Flusser, 2011: 6)—such as pixels, photons, bit and bites, or data points—produced by an apparatus (Flusser, 2000: 14). For a technical image to become possible, we need the detour of scientific texts: texts abstract from images, and apparatuses produce technical images as abstractions from texts (Figure 2).

An example of a technical image produced by an apparatus is the scan of an organ that a radiologist examines. The scan is not a depiction of that organ (and perhaps the cancer that feeds on it), but the visualization of the scattering of photons on a light-sensitive plate. A diagnostic AI system trained in image ⌈recognition⌉ does not ⌈recognize⌉ the organ (and perhaps the tumor); it compares the pattern of particles with those on many other such images and calculates their correspondences. A simplified example is offered in Figure 3. Here, the AI system ⌈recognizes⌉ a hand-written number "8," a *traditional* image (Figure 3a) after it has been digitized as a collection of pixels of varying intensity in a grid (Figure 3b)—that is, as a *technical* image.[9] Thus, there is an "epistemological gap" (Newlands, 2021; cf. Smith, 2019) between what the radiologist is able to observe in the patient's organ and the

[8] For rhetorical reasons—the simile of the technical *image*—as well as in the interest of parsimony, we mostly draw our discussion and examples from the particular AI functionality of image ⌈recognition⌉. It should be noted, however, that the same arguments can be made regarding other functionalities of current AI, including text ⌈recognition⌉ and reinforcement ⌈learning⌉.

[9] In an interesting section, "Fooling Deep Neural Networks," Mitchell (2019: 128ff) discusses research that shows how subtle changes in the pixel structure may not make a difference to our recognition of the represented image but completely distort the image ⌈recognition⌉ by the AI system.

FIGURE 3
Traditional Image versus Technical Image



A handwritten number "8" (panel A) rendered into a technical image (panel B) after its digitalization into an 18 × 18 resolution grid. The AI image ⌈recognition⌉ works by finding patterns in pixel intensity across a large number of such technical images both exhibiting number "8" and not number "8." Pixel intensity is a "quantum"; Flusser would refer to these pixels as "particles" (image source: Mitchell, 2019, from Figures 2 and 3).

technical image that is produced by the apparatus she operates, and between the hand-written number "8" and its reproduction by the AI system.

Viana (2018: 80) discussed this epistemological gap as there being two layers in technical images:

> On one layer, the most complex and detailed part of the representation is accessible only to the apparatus at its inception, while, on the other, the human viewer receives a surface that, to a great extent, does not differ from what she [the human viewer] is already used to experiencing with traditional images or texts.

We, human viewers, tend to see and perceive the technical image as a traditional image, whereas its technicality is black-boxed, hidden "under the hood" of the AI. Without recognizing this double layer, we are at risk of "mistaking the map for the territory." Smith (2019) explained how AI systems, like human beings, make registrations of the world to abstract from its detail, and, in doing so, "approximate, do violence, privilege some things at the expense of others" (Smith, 2019: 111). What matters, here, is not the registration itself, but "*that which* is registered" (Smith, 2019: 112, emphasis added) and *how* it is registered. That is, in the technical image the semantic relationship has been severed; the AI system is impartial—it has no interest in, does not know about, cannot make sense of, and is not committed—to *that which* it registers. "What limits [AI systems] is that, so far, nothing matters *to them*. To use a phrase of which Haugeland was fond: *they don't give a damn*" (Smith, 2019: 108, original emphasis).

This detachment of AI systems from the world is problematic, if we do not recognize or ignore it (such as in scenario B). Then, the technical image may ⌈goad⌉ or ⌈lure⌉ us into adopting and accommodating the formal rationality of the discursive structure of its deeper, more complex yet hidden layer: this is ontological assimilation. As we have shown above, this process is already happening. We already lose our sense of judgment, and un-learn our ability to socially engage in judgment. We increasingly rely on apparatuses and their technical images to inform, guide, and steer our judgment; we increasingly learn to be helpless (Moore, 2019). We are in scenario C while believing that we are in scenario B, such that we risk ending up in scenario A; this engagement with AI systems makes us to serve their reckoning.

For Flusser, ending in scenario A is not the necessary outcome of our engaging with AI systems and their technical images (Flusser, 2011: 79ff). According to him, we can still use them as means to creating new information, which may then inform judgment and thereby be meaningful for decision-making, but only if we understand the nature of the technical images that AI systems produce and their difference from traditional images. We pick up on this possibility in the closing section.

DISCUSSION AND CALL FOR (IN)ACTION

Our article offered the provocation that algorithmic ⌈morality⌉ is to be avoided and resisted if we want to maintain a morality that relies on judgment. For this reason, we see this article as an act of "disciplined provocation" (Vince & Hibbert, 2018) at the theoretical level, and as a political intervention (Gabriel, 2016) at the practical level. In what follows, we first consolidate the theoretical insights generated in this article, and foreshadow their implications for future theorizing and empirical endeavors.

In line with this journal's focus on actionable research (Bartunek & Egri, 2012), we close with calls for both *in*action and action.

We argued that we need to reconsider how we develop, understand, and apply algorithms in our daily lives and businesses. In particular, we problematized the ontological assumptions underlying human judgment and algorithmic reckoning. In doing so, we argued that we need to acknowledge the possibility of ontological assimilation (unrecognized in scenario B), because it enables us to recognize and deal with the current situation of co-constituted morality (scenario C). AI systems *do* have agency, which—when unrecognized and unchecked—enables them to inform, guide, and steer human judgment in decision-making (scenario A). From this perspective, algorithms are not external to our morality (scenario C) and, therefore, they cannot merely be used as innocent tools in decision-making (scenario B). Wise and prudent usage of AI systems depends on understanding how they are—in the language of Flusser—apparatuses that produce technical images.

What difference would it make to acknowledge that we are in scenario C? We offer suggestions concerning the acknowledgment of AI agency, the limits of technical images, and AI developers' motivations. First, both the thought that AI can be made to serve human needs and the ever-growing number of instances of indiscriminate use of AI in all sorts of managerial tasks (and other instances of decision-making) are expressions of hubris, an arrogant attitude espousing over-confidence and contempt for the advice and criticism of others. In the former case, because of the erroneous belief that full control over AI is possible; in the latter case, because of the inability or unwillingness to understand how AI advances formal rationality through digitalization (e.g., this paper; Flyverbom, 2019; Lindebaum et al., 2020; Smith, 2019). Such hubris may have potentially destructive outcomes (Sadler-Smith & Cojuharenco, 2021). For example, lack of acknowledgment of the influence of AI on organizational learning is likely to enhance the "myopia of learning" in organizations (Balasubramanian et al., 2020).

Second, we should start to learn using technical images as models, rather than as representations or maps (Gabriel, 2018). Literally so. We offer a pretty mundane example of using a technical image as a map, one that would be amusing had it not been so tragic. A Dutch Broadcasting Foundation news item from last year revealed that several people had driven their cars into the harbor of Marseille, France (Nederlandse Omroep Stichting, 2020). Apparently, the drivers blindly followed the instructions from their navigation systems. However, and unfortunately enough for them, the navigation system was not up to date and led them to drive into the water. This example is a stand-in for a more general tendency to interpret AI's technical images as if they were pieces of ⌈intelligent⌉ advice that correctly represent what the road ahead looks like.

Third, in terms of developing AI, we can better understand some of the motivations that propel software developers in pushing the limits of AI. In line with Flusser (2000), AI systems are "play-things": they challenge software developers to find novel possibilities with and for the technology, which then leads to new, "improved" AI systems the prowess of which exceeds the limitations of the previously available AI systems (first, "beat Atari," then "beat chess," then "beat Go"). Acknowledging the relevance of scenario C would enable us think of such software development as an impressive accomplishment in programming—and not as an inevitable solution to an as yet unknown problem.

Beyond these points, and more fundamentally, a co-constitutive perspective allows us to get to terms with the expectation that, in all likelihood, "the existential interests of future men and women will focus on technical images" (Flusser, 2011: 4). Artificially intelligent algorithms and the technical images they produce are here to stay, so the question is how to live with them. In this respect, Flusser (2011: 4) sketched "two opposing possibilities for the post-historical society of technical images." One possibility is "negative," a dystopia in which artificially intelligent algorithms are the backbone of a totalitarian society and in which "human beings operate as a function of the apparatus. A man gives an apparatus instructions that the apparatus has instructed him to give" (Flusser, 2011: 74), such that we end up living a life that is a function of AI (i.e., scenario A). This dystopia is already happening to some extent. For example, Newlands (2021) vividly described how workers in the gig economy are being controlled by algorithmic data processing. For algorithmic decision-making in the gig economy, the data feeding the algorithm has to be "collectable in a format that can be read and understood by the algorithm" (Newlands, 2021: 11). What happens here is the assimilation of activity to data points, which corresponds to our analysis above. However, there is nothing inevitable in this dystopia. The time is *now* to become aware of the challenges, opportunities, and dangers that we have created. The time is *now* to reflect on what we are actually doing with AI. Hence, a call for *in*action when it comes to the indiscriminate, unreflective use of AI systems in yet other situations, and a call for action in rethinking how we would want to use the technology given its affordances and limitations.

In light of this, we recall that, more than a quarter century ago, there was a vehement discussion about the uncertainties and risks of the then-novel technology of genetic modification of biological organisms (GMO). Back then, a plea for honoring the precautionary principle was often heard (e.g., Andorno, 2004; O'Riordan & Cameron, 1995). The precautionary principle stated that it is wise to stop developing a technology if there is a risk of that technology precipitating fundamental, irreversible change. With current AI, we are in a similar situation as with GMO technology a quarter century ago. However, instead of applying the precautionary principle with AI, and particularly in a context of learning and education, we witness that the reverse is happening: AI is being developed at lightning speed. In the realm of management education, this includes automated feedback, digital assistants, and virtual reality applications (Chace, 2020; Lewis, 2013). Although there is some sensitivity in the public policy domain (e.g., European Commission, 2020) for possible problems with AI, this is, as we have hopefully convinced the reader by now, not even close to appreciating the dangers that AI brings with it. We find ourselves in a situation in which developing and using the technology can, and already does, lead to fundamental and irreversible changes, as AI already infiltrates our daily life on almost every dimension. The above dual call for (*in*)action amounts to invoking the precautionary principle, such that we use AI systems for reckoning tasks, and not for judgment, which is beyond their capacity (cf. Smith, 2019).

In this way, we can imagine what Flusser's (2011) other, "positive" possibility may entail: a future of democracy and freedom that is *supported* by our intelligent use of AI. This second possibility is embedded in scenario C. It suggests that we may preserve the informational and decision-support function of technical images when they serve this function as a resource for learning in (managerial) deliberation (cf. Gersel & Johnsen, 2020). However, we can only do so on the condition that we duly recognize their hidden, formally rational discursive structure. We are not just already in scenario C, but staying there will demand a lot of effort and constructive-critical thinking in a pragmatist style to keep ontological assimilation at bay. Hence a call for *action*.

For example, Berti, Nikolova, Jarvis, and Pitsis (2021) reminded us of the importance of educators' and students' rich understanding of ethical challenges. As AI is central to so many organizations and organizational processes, it stands to reason that it becomes part of the fabric of "ambiguities, unforeseen consequences, paradoxes and contrasting interests" that complicate managerial practice (Berti et al., 2021: 3). Indeed, business ethics learning should be informed by judgment (Berti et al., 2021) and be taught in courses that are spread across the curriculum (Parks-Leduc, Mulligan, & Rutherford, 2021). After all, teaching business ethics should be about encouraging moral awareness and imagination (Hartmann, 2006) by confronting students with questions critical of prevailing business practices and received wisdom. Our article provides teachers and learners of business ethics with the heuristics to do just that in the context of AI: question taken-for-granted and often implicit assumptions about decision-making and other organizational processes informed by AI.

To conclude, our call to (in)action is a means to embrace again efforts to retain, or maybe resurrect, a sense of morality and judgment that is under threat of AI reckoning in decision-making. This is crucial in the context of management education, in business practices, and beyond. Given the rapid speed of AI development, we need to get to grips with the "spirits that we cited" and learn anew how to make decisions that are informed by our own judgment rather than by algorithmic reckoning.

REFERENCES

Academy of Management. 2021, March 17. Academy of Management: Style guide for authors. Retrieved from https://aom.org/docs/default-source/publishing-with-aom/aom_journal_style_guidea3b84b773e3649569a17a05e14cc6eaf.pdf

Agamben, G. 2009. ***What is an apparatus?*** Stanford, CA: Stanford University Press.

Alonso, E. 2014. Actions and agents. In K. Frankish & W. M. Ramsey (Eds.), ***The Cambridge handbook of artificial intelligence***: 232–246. Cambridge, MA: Cambridge University Press.

Ames, M. G. 2018. Deconstructing the algorithmic sublime. ***Big Data & Society***, 5: 1–4.

Anderson, M., & Anderson, S. L. (Eds.). 2011. ***Machine ethics***. Cambridge, U.K.: Cambridge University Press.

Andorno, R. 2004. The precautionary principle. ***Journal of International Biotechnology Law***, 1: 11–19.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. 2018. The Moral Machine experiment. ***Nature***, 563: 59–64.

Azelvandre, J. P. 2001. Constructing sympathy's forge. ***Philosophy of Education***, 2001: 170–178.

Bachrach, P., & Baratz, M. S. 1963. Decisions and nondecisions. ***American Political Science Review***, 57: 632–642.

Balasubramanian, N., Ye, Y., & Xu, M. 2020. Substituting human decision-making with machine learning: Implications for organizational learning. ***Academy of Management Review***, 20. doi: 10.5465/amr.2019.0470

Balconi, M., Fronda, G., Venturella, I., & Crivelli, D. 2017. Conscious, pre-conscious and unconscious mechanisms in emotional behaviour. ***Applied Sciences***, 7: 1280–1293.

Bartunek, J. M., & Egri, C. P. 2012. Can academic research be managerially actionable? What are the requirements for determining this? ***Academy of Management Learning & Education***, 11: 244–246.

Berti, M., Nikolova, N., Jarvis, W., & Pitsis, A. 2021. Embodied phronetic pedagogy: Cultivating ethical and moral capabilities in postgraduate business students. ***Academy of Management Learning & Education***, 20: 6–29.

Bijker, W. E., Hughes, T. P. & Pinch, T. F. (Eds.). 1987. ***The social construction of technological systems***. Cambridge, MA: MIT Press.

Bonfert, M., Spliethöver, M., Arzaroli, R., Lange, M., Hanci, M., & Porzel, R. 2018. If you ask nicely. ***Proceedings of the 2018 on International Conference on Multimodal Interaction/ICMI18***, doi: 10.1145/3242969.3242995.

Broussard, M. 2018. ***Artificial unintelligence***. Cambridge, MA: MIT Press.

Brunsson, K., & Brunsson, N. 2017. ***Decisions***. Cheltenham, U.K.: Edward Elgar.

Bryson, J. J. 2018. Patiency is not a virtue. ***Ethics and Information Technology***, 20: 15–26.

Chace, C. 2020, October 29. The impact of artificial intelligence on education. ***Forbes***. Retrieved from https://www.forbes.com/sites/calumchace/2020/10/29/the-impact-of-artificial-intelligence-on-education

Cohen, M. D., March, J. G., & Olsen, J. P. 1972. A garbage can model of organizational choice. ***Administrative Science Quarterly***, 17: 1–25.

Dewey, J. 1897. ***My pedagogic creed***. New York, NY: E. L. Kellogg & Co.

Dewey, J. 1916. ***Essays in experimental logic***. Chicago, IL: University of Chicago Press.

Dewey, J. 1922. ***Human nature and conduct***. New York, NY: Henry Holt.

Dewey, J. 1929. ***The quest for certainty***. New York, NY: Minton, Balch.

Dewey, J. 1939. ***Theory of valuation***. Chicago, IL: Chicago University Press.

Dewey, J., & Tufts, J. H. 1932. ***Ethics***. New York, NY: Henry Holt.

Dignum, V. 2019. ***Responsible artificial intelligence***. Cham, Switzerland: Springer.

European Commission. 2020. ***On artificial intelligence: A European approach to excellence and trust*** (White paper COM/2020/65). Retrieved from https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Elmholdt, K. T., Elmholdt, C., & Haahr, L. 2021. Counting sleep: Ambiguity, aspirational control and the politics of digital self-tracking at work. ***Organization***, 28: 164–185.

Etzioni, A., & Etzioni, O. 2017. Incorporating ethics into artificial intelligence. ***Journal of Ethics***, 21: 403–418.

Floridi, L. 2019. Establishing the rules for building trustworthy AI. ***Nature Machine Intelligence***, 1: 261–262.

Floridi, L., & Cowls, J. 2019. A unified framework of five principles for AI in society. ***Harvard Data Science Review***, 1. doi: 10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. 2018. AI4People: An ethical framework for a good AI society. ***Minds and Machines***, 28: 687–707.

Flusser, V. 2000. ***Towards a philosophy of photography***. London, U.K.: Reaktion Books.

Flusser, V. 2011. ***Into the universe of technical images***. Minneapolis, MN: Minnesota University Press.

Flyverbom, M. 2019. ***The digital prism***. Cambridge, U.K.: Cambridge University Press.

Foucault, M. 2008. ***The birth of biopolitics***. London, U.K.: Palgrave MacMillan.

Gabriel, M. 2018. ***Der Sinn des Denkens***. Berlin, Germany: Ullstein.

Gabriel, Y. 2016. The essay as an endangered species: Should we care? ***Journal of Management Studies***, 53: 244–249.

Gadamer, H.-G. 1996. ***The enigma of health***. Stanford, CA: Stanford University Press.

Gersel, J., & Johnsen, R. 2020. Towards a novel theory of rational managerial deliberation: Stakeholders, ethical values, and corporate governance. ***Academy of Management Learning & Education***, 19: 269–288.

Gibson, J. 1979. ***The ecological approach to visual perception***. Reading, MA: Houghton Mifflin.

Glaser, V. L., Pollock, N., & D'Adderio, L. 2021. The biography of an algorithm: Performing algorithmic technologies in organizations. ***Organization Theory***, 4. doi: 10.1177/26317877211004609

Glikson, E., & Woolley, A. W. 2020. Human trust in artificial intelligence: Review of empirical research. ***Academy of Management Annals***, 14: 627–660.

Greene, J., & Haidt, J. 2002. How (and where) does moral judgment work? ***Trends in Cognitive Sciences***, 6: 517–523.

Habermas, J. 1993. ***Justification and application***. Cambridge, MA: MIT Press.

Habermas, J. 1996. ***Between facts and norms***. Cambridge, MA: MIT Press.

Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. ***Psychological Review***, 108: 814–834.

Hartmann, E. M. 2006. Can we teach character? ***Academy of Management Learning & Education***, 5: 68–81.

Heidegger, M. 1977. ***The question concerning technology***. New York, NY: Garland.

Helin, S., Jensen, T., Sandstrom, J., & Clegg, S. R. 2011. On the dark side of codes. ***Scandinavian Journal of Management***, 27: 24–33.

Henley, J. 2021, January 15. Dutch government resigns over child benefits scandal. ***Guardian***. Retrieved from https://www.theguardian.com/world/2021/jan/15/dutch-government-resigns-over-child-benefits-scandal

Hibbert, P., & Cunliffe, A. 2015. Responsible management: Engaging moral reflexive practice through threshold concepts. ***Journal of Business Ethics***, 127: 177–188.

Introna, L. D. 2014. Towards a post-human intra-actional account of sociomaterial agency (and morality). In P. Kroes & P. P. Verbeek (Eds.), ***The moral status of technical artefacts***: 31–54. Dordrecht, The Netherlands: Springer.

Jobin, A., Ienca, M., & Vayena, E. 2019. The global landscape of AI ethics guidelines. ***Nature Machine Intelligence***, 1: 389–399.

Johnson, D. G. 2006. Computer systems: Moral entities but not moral agents. ***Ethics and Information Technology***, 8: 195–204.

Kellogg, K. C., Valentine, M. A., & Christin, A. 2020. Algorithms at work: The new contested terrain of control. ***Academy of Management Annals***, 14: 366–410.

Kolb, A. Y., & Kolb, D. A. 2005. Learning styles and learning spaces: Enhancing experiential learning in higher education. ***Academy of Management Learning & Education***, 4: 193–212.

Kolb, A. Y., & Kolb, D. A. 2009. Experiential learning theory. In S. Armstrong & C. Fukami (Eds.), ***The SAGE handbook of management learning, education and development***: 42–68. Los Angeles, CA: SAGE.

Lawson, T. 2019. ***The nature of social reality***. New York, NY: Routledge.

Latour, B. 2005. ***Reassembling the social***. Oxford, U.K.: Oxford University Press.

Leonardi, P. M. 2012. Materiality, sociomateriality, and socio-technical systems. In P. M. Leonardi, B. A. Nardi, & J. Kallinikos (Eds.), ***Materiality and organizing***: 25–48. Oxford, U.K.: Oxford University Press.

Lewis, J. K. 2013. ***Ethical implementation of an automated essay scoring (AES) system*** (Faculty and Staff Articles & Papers, Paper 47). Retrieved from http://digitalcommons.salve.edu/fac_staff_pub/47

Lindebaum, D., Geddes, D., & Gabriel, Y. 2017. Moral emotions and ethics in organisations. ***Journal of Business Ethics***, 141: 645–656.

Lindebaum, D., Vesa, M., & den Hond, F. 2020. Insights from The Machine Stops to better understand rational assumptions in algorithmic decision making and its implications for organizations. ***Academy of Management Review***, 45: 247–263.

Loon, M. 2021. Practices for learning in early careers. ***Academy of Management Learning & Education***, 20: 182–202.

MacCormick, J. 2012. ***Nine algorithms that changed the future***. Princeton, NJ: Princeton University Press.

March, J. G. 1994. ***A primer on decision making***. New York, NY: Free Press.

Martela, F. 2015. Fallible inquiry with ethical ends-in-view: A pragmatist philosophy of science for organizational research. ***Organization Studies***, 36: 537–563.

Martin, K. E. 2019. Designing ethical algorithms. ***MIS Quarterly Executive***, 18: 129–142.

Mitchell, M. 2019. ***Artificial intelligence. A guide for thinking humans***. London, U.K.: Pelican.

Moore, P. V. 2019. Book review—Artificial intelligence: What everyone needs to know. ***Organization Studies***, 40: 466–470.

Moser, C., Reinecke, J., den Hond, F., Svejenova, S. V., & Croidieu, G. 2021. Biomateriality and organizing: Towards an organizational perspective on food. ***Organization Studies***, 42: 175–193.

Murray, A., Rhymer, J., & Sirmonet, D. G. 2021. Humans and technology: Forms of conjoined agency in organizations. ***Academy of Management Review***, 46: 552–571.

Nederlandse Omroep Stichting. 2020, March 3. Fout navigatiesysteem leidt auto's naar water in haven Marseille [Faulty navigation system leads cars to water in Marseille port] (in Dutch). Retrieved from https://nos.

nl/artikel/2325654-fout-navigatiesysteem-leidt-auto-s-naar-water-in-haven-marseille

Newlands, G. 2021. Algorithmic surveillance in the gig economy: The organisation of work through Lefebvrian conceived space. ***Organization Studies***, 42: 719–737.

Nyberg, D. 2008. The morality of everyday activities: Not the right, but the good thing to do. ***Journal of Business Ethics***, 81: 587–598.

O'Neill, C. 2016. ***Weapons of math destruction: How big data increases inequality and threatens democracy***. New York, NY: Crown Publishers.

O'Riordan, T., & Cameron, J. (Eds.). 1995. ***Interpreting the precautionary principle***. London, U.K.: Earthscan.

Orlikowski, W. J. 2007. Sociomaterial practices: Exploring technology at work. ***Organization Studies***, 28: 1435–1448.

Parks-Leduc, L., Mulligan, L., & Rutherford, M. A. 2021. Can ethics be taught? Examining the impact of distributed ethical training and individual characteristics on ethical decision making. ***Academy of Management Learning & Education***, 20: 30–49.

Pereiro, L., & Lopes, A. B. 2020. ***Machine ethics: From machine morals to the machinery of morality***. Cham, Switzerland: Springer.

Putnam, H. 2002. ***The collapse of the fact/value dichotomy and other essays***. Cambridge, MA: Harvard University Press.

Redden, J., Brand, J., & Terzieva, V. 2020, August. Data harm record (updated). Retrieved https://datajusticelab.org/data-harm-record

Raisch, S., & Krakowski, S. 2021. Artificial intelligence and management: The automation–augmentation paradox. ***Academy of Management Review***, 46: 192–210.

Sadler-Smith, E., & Cojuharenco, I. 2021. Business schools and hubris: Cause or cure? ***Academy of Management Learning & Education***, 20: 270–289.

Schumann, F. 2020, March 21. We have to bring down the number of cases now. Otherwise we won't be able to handle it. ***Zeit***. Retrieved from https://www.zeit.de/wissen/gesundheit/2020-03/christian-drosten-coronavirus-pandemic-germany-virologist-charite

Shotter, J., & Tsoukas, H. 2014. In search of phronesis: Leadership and the art of judgment. ***Academy of Management Learning & Education***, 13: 224–243.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. 2017. Mastering the game of Go without human knowledge. ***Nature***, 550: 354–359.

Simpson, B., & den Hond, F. 2021. The contemporary resonances of classical Pragmatism for studying organization and organizing. ***Organization Studies***. doi: 10.1177/0170840621991689

Smith, B. C. 2019. ***The promise of artificial intelligence***. Cambridge, MA: MIT Press.

Sun, R. 2014. Connectionism and neural networks. In K. Frankish & W. M. Ramsey (Eds.), ***The Cambridge handbook of artificial intelligence***: 108–127. Cambridge, U.K.: Cambridge University Press.

Tegmark, M. 2017. ***Life 3.0: Being human in the age of artificial intelligence***. New York, NY: Alfred A. Knopf.

Two Minute Papers. 2015, March 7. ***Google DeepMind's Deep Q-learning playing Atari Breakout*** [Video file]. Retrieved from https://www.youtube.com/watch?v=V1eYniJ0Rnk

Vesa, M., & Tienari, J. 2020. Artificial intelligence and rationalized unaccountability: Ideology of the elites? ***Organization***. doi: 10.1177/1350508420963872

Viana, D. 2018. Two technical images: Blockchain and high-frequency trading. ***Philosophy & Technology***, 31: 77–102.

Vince, R., & Hibbert, P. 2018. Disciplined provocation: Writing essays for AMLE. ***Academy of Management Learning & Education***, 17: 397–400.

Wallach, W., & Allen, C. 2009. ***Moral machines***. Oxford, U.K.: Oxford University Press.

Watson, T. J. 2013. Pragmatism, organizations and getting to grips with reality. In M. Kelemen & N. Rumens (Eds.), ***American pragmatism and organisation***: 59–72. Farnham, U.K.: Ashgate.

Weber, M. 1968. ***Economy and society***. Berkeley, CA: University of California Press.

Wong, D. B. 2006. ***Natural moralities***. Oxford, U.K.: Oxford University Press.

Zhao, K., Stylianou, A. C., & Zheng, Y. 2018. Sources and impacts of social influence from online anonymous user reviews. ***Information & Management***, 55: 16–30.

Zuboff, S. 1988. ***In the age of the smart machine***. New York, NY: Basic Books.

Zuboff, S. 2019. ***The age of surveillance capitalism***. New York, NY: Public Affairs.

Christine Moser (c.moser@vu.nl) is Associate Professor of Organization Theory at Vrije Universiteit Amsterdam. She conducts research on corporate social responsibility, knowledge flows in social networks, and the role of technology in social interaction. Christine has published in, among others, Research Policy, Human Relations, Organization Studies, New Media and Society, and Business & Society.

Frank den Hond (frank.denhond@hanken.fi) is the Ehrnrooth professor of management and organization at Hanken School of Economics, Helsinki, Finland, and a past editor in chief of *Organization Studies*. His work has been published in many of the top journals in management and organization. As per August 2021, he serves as coeditor in chief of *Business Ethics Quarterly*.

Dirk Lindebaum is senior professor in organisation & management. He studies the mechanisms through which freedom is lost and (re)gained at work, be it through emotional or technological means. His articles & essays have been consistently published in journals of international distinction. Currently, he serves as associate editor for the *Academy of Management Learning & Education* (essay section). The relevance of his work is regularly recognized in news outlets (visit https://dirklindebaum.EU).