This is a postprint version. Please cite as: Buhmann, A., & Fieseler, C. (2021). Tackling the Grand Challenge of Algorithmic Opacity Through Principled Robust Action. Morals & Machines, 1(1), 74-85.

Tackling the Grand Challenge of Algorithmic Opacity Through Principled Robust Action

ABSTRACT:
Organizations increasingly delegate agency to artificial intelligence. However, such systems can yield unintended negative effects as they may produce biases against users or reinforce social injustices. What pronounces them as a unique grand challenge, however, are not their potentially problematic outcomes but their fluid design. Machine learning algorithms are continuously evolving; as a result, their functioning frequently remains opaque to humans. In this article, we apply recent work on tackling grand challenges though robust action to assess the potential and obstacles of managing the challenge of algorithmic opacity. We stress that although this approach is fruitful, it can be gainfully complemented by a discussion regarding the accountability and legitimacy of solutions. In our discussion, we extend the robust action approach by linking it to a set of principles that can serve to evaluate organisational approaches of tackling grand challenges with respect to their ability to foster accountable outcomes under the intricate conditions of algorithmic opacity.

The proliferation of artificial intelligence (AI) in business, public administration, and everyday life has emerged as a grand challenge that calls for coordinated action (George,Howard-Grenville,Joshi,&Tihanyi, 2016). Since the early 2000s, we have witnessed a re-emergence in interest regarding the widespread implementation of autonomous technologies after decades of modest advances in the development of AI in the late 1960s, when military and other institutional funding for AI research had levelled. This AI renaissance is fuelled by new opportunities for ubiquitous digital data collection and revolutionary automated learning methods that made earlier promises of novel prediction- and decision-making approaches a reality. Between the early 2000s and today, the societal relevance of AI rose dramatically: self-learning systems based on algorithms are not only used to make sense of enormous amounts of existing data, but can also be used to make predictions about the future. As such, they are fast becoming crucial tools for organizational management (Hildebrandt, 2008; Kim et al., 2014).

Machine learning systems have already taken over roles in reshaping work, as well as in defence and policy making, but also by helping to solve existing grand challenges (e.g. ecological degradation, disease treatments or fossil energy dependency). While organizations make extensive use of algorithms as agents of complex computerised decision-making, the input data they use can be biased, their deep learning operations are often invisible, and their recommendations and decisions yield often unintentional negative effects (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016). Therefore, they may yield negative consequences, e.g., for equality, privacy, stock and commodity exchange or democratic election outcomes (Barnet, 2009; Tutt, 2016; Zarsky, 2016). At the same time, as these machine learning systems are often, by design, complex and continuously evolving, their functioning frequently remains opaque to humans and the decisions they make are often implicit and invisible (Beer, 2009; Pasquale, 2015). Indeed, the proliferation of opaque algorithms significantly challenges established procedures of maintaining accountability and legitimacy (Martin, 2018; Buhmann, Paßmann, & Fieseler, 2020).

Grand challenges are marked by complexity with often unknown or conflicting solutions and technical and social elements that are intertwined. They involve circular causality, an absence of well-structured alternative solutions, numerous interactions and associations, emergent understandings and nonlinear dynamics, and result in organizations facing radical uncertainty (Martí, 2018). In recent work on grand challenges, scholars have underscored the potential of "robust strategies" (of participation, multivocal inscription, and distributed experimentation) for the generation of novel solutions and sustained engagement (Ferraro, Etzion, & Gehman, 2015; Etzion, Gehman, Ferraro, & Avidan, 2017). Focal actors facilitate participation and multivocality to enable a form of experimental "free play" that works toward novelty and creativity of solutions. But how do these actors stay accountable in the process and what makes the outcomes both novel and legitimate?

In this paper we apply the robust action approach to discuss strategies for tackling the grand challenge of algorithmic opacity. Further, when focusing on issues that pertain to accountability and outcome legitimacy, we argue in favour of amending the approach with a set of communicative principles that substantiate a critical point of view to assess the extent to which applied strategies may allow for accountability under conditions of algorithmic opacity and for the generation of legitimate outcomes. The paper aims to contribute to extant literature in three main ways. First, we add to recent work on grand challenges (George et al., 2016; Etzion et al., 2017; Dentoni, Bitzer, & Schouten, 2018) by introducing algorithmic opacity as a grand challenge, specifically pointing to the technical and procedural aspects that call for participative "robust action" approaches for generating novel solutions. Second, we add to literature on the opacity and ethics of algorithms (Ananny, 2016; Burrell, 2016; Martin, 2018; Mittelstadt et al., 2016) on the one hand and stakeholder engagement and accountability (Palazzo & Scherer, 2006; Seele & Lock, 2015; Gilbert & Rasche, 2007; Rasche & Esser, 2006; Scherer, Palazzo, & Seidl, 2013) on the other, as we identify challenges in which negotiation parties are burdened by the fluidity and poor transparency of self-learning systems, which leads us to argue for communicative principles (cf. Buhmann et al., 2020) to assess the accountability and legitimacy of novel solutions. Finally, we specifically suggest the aforementioned notion of communicative principles as a normative extension to the current pragmatist approach to tackling grand challenges (Ferraro et al. 2015).

THE GRAND CHALLENGE OF ALGORITHMIC OPACITY

Opacity of machine learning algorithms as a grand challenge

In contrast to the global proliferation and societal penetration of earlier technologies, such as the car, electricity or the telephone, modern algorithmic decision systems come with a special kind of opacity: Machine learning algorithms are not a set of rules defined by programmers, but by algorithmically produced rules of learning: "The internal decision logic of the algorithm is altered as it 'learns' on training data" (Burrell, 2016, p. 5). Plainly put, algorithms are used to program new algorithms. In many cases, their outcomes cannot be observed in the 'laboratory' of software engineering, but only in the 'field' of actual usage by different user groups over long temporal periods (e.g. when the learning data for machine learning algorithms are produced by actual users over many years – such as in almost every algorithmic decision system attributed as AI).

As a result, the core societal issue with algorithmic decision systems—on the one hand—is that they cannot usually be accessed for public scrutiny, as they are proprietary entities of the organisations that own or license them. In this case, they elude access for strategic reasons, such as to ensure functionality, competitiveness or the confidentiality of data (Ananny & Crawford, 2016; Glenn & Monteith, 2014; Leese, 2014; Stark & Fins, 2013). On the other hand, it seems to be increasingly important that they elude access for technical and procedural reasons: First, they are based, in part, on structurally inaccessible and incomprehensible procedures—not simply to the public, but also to the organisations that own and employ them, and even to specialists (Ananny, 2016; Burrell, 2016). Second, they are highly fluid technologies that evolve only in the 'field' (Sandvig et al, 2016).

According to Ferraro et al. (2015), grand challenges are characterised by: (1) complexity (i.e. that the process has many different and heterogeneous actors with emergent understandings); (2) a radical uncertainty, which means—put bluntly—that participants are not easily able to foresee future consequences of their current actions; and (3) a strong evaluativeness, meaning that the values and valuations of actions are, to a significantly visible extent, not clear from the beginning, but are being produced within a longer process of production and co-production of meaning. The authors stress that in such constellations, it is more helpful to frame processes as a collective efforts, "rather than the achievement of a single organization" (ibid., p. 2). That means that within the core idea of tackling grand challenges pragmatically, in these extremely collective processes of knowledge production, it is more important than ever to stress the cooperativeness of interaction. The complexity of grand challenges is especially produced by "the large array of actors involved, and the manner in which they associate and interact" (Ferraro et al., 2015, p. 3) They also are seemingly intractable, resisting easy fixes" (ibid., p. 3). This extension happens in a double sense: It affects people beyond the immediate reach of relevant organisations. The organisations are also in need of assessment beyond their own boundaries. Grand challenges typically unfold, not within a single organisation, but at the field level, where actors and actions are more distributed, diverse and more difficult to govern than they are within organisations. Any understanding of a shared issue is likely to be continuously (re)negotiated (Grodal & O'Mahony, 2017). Understood as such, algorithmic opacity cannot simply be 'tackled' by demanding organisations to 'make their algorithms transparent.' There is no straightforward way to address poorly transparent and highly fluid algorithmic processes and organisations cannot simply deliver accounts for these technologies (Buhmann et al., 2020). They need to be addressed in a participative and discursive process together with their stakeholders; they need the 'pragmatic treatment' that Ferraro et al. (2015) proposed for other grand challenges.

TACKLING ALGORITHMIC OPACITY THROUGH ROBUST ACTION STRATEGIES

Treating grand challenges pragmatically through robust action entails addressing them through participatory architectures (the adoption of a structural dimension, including forums of participation through which concerned and affected actors may interact and debate with each other), multivocal inscriptions (instantiation of debate and discourse, and the inscription of differing viewpoints in material forms), and distributed experimentation (a practice dimension of investigation and testing that can point to solutions that might work and identify any that do not) (cf. Ferraro et al. 2015). In the following section, we will apply this framework with a specific focus on algorithmic opacity and accountability. In doing so, we will stress that the quality of this process is shaped by the

specific way in which organisations engage with the emergent demands of their stakeholders (Greenwood, Raynard, Kodeih, Micelotta, & Lounsbury, 2011). Accordingly, engagement—as a practice undertaken by organisations to involve stakeholders – has been described as a way to achieve accountability (Gray, 2002; Van Buren, 2001), particularly for AI and algorithmic systems (Buhmann et al., 2020): In the case of self-learning algorithms, where external demands are often unclear and/or no clear-cut accountability standards are available, organisations need to engage with their various stakeholders to create such standards. However, more engagement does not automatically mean more accountability; while some engagement practices may indeed be focussed on deliberation, listening, and learning (Edwards, 2016; Romenti 2010), others may primarily aim at creating an image of accountability (Swift, 2001) or even be outright deceptive (Greenwood, 2007).

In this section, we employ the robust action framework, as proposed by Ferraro et al. (2015), to showcase the applicability and the boundary cases for issues of algorithmic opacity and accountability. In particular, we want to point to the critical importance to establish a level playing field with informed actors to establish legitimate robust action in the first place, an—in our view—important point that is yet underdeveloped in the pragmatist framework. To this end, we want to illustrate our analysis with two case studies of algorithmic opacity, namely content personalisation systems and autonomous vehicles.

With regard to content personalisation systems (CPSs), consider that while traditional media broadcast content to large heterogeneous audiences, most people today receive highly personalised content through social media, search engines, and targeted advertisements (Bucher, 2012; Goldman, 2006). This happens based on systems that curate tailored information to individual users "through interactions of (a) prioritization algorithms that decide which topics are (and are not) trending (...); (b) profiling algorithms that infer user preferences and attributes from small patterns or correlations, by which individuals are clustered into meaningful groups according to their behavior, preferences, and other characteristics (...); and (c) automated bots that post and interact directly with users to promote certain content or viewpoints" (Mittelstadt, 2016, pp. 4991). CPSs on popular platforms such as Twitter, Facebook, and Reddit have been shown to significantly interfere with politics (Woolley, 2016), such as by segmenting audiences of like-minded people into highly self-reinforcing networks ("echo chambers") (Leese, 2014). Thus, CPSs have the potential to undermine open exchanges of ideas in political debate. This possibility has raised significant public concern and led to calls for system transparency. Furthermore, it has raised strong ethical concerns regarding the duties of service providers that work with CPSs (Mittelstadt, 2016).

Second, we must cons der that autonomous vehicles come equipped with a variety of sensors that obtain data and information from the environment to serve as input for software that guides the vehicles through traffic (Bagloee et al., 2016). Several safety and ethical concerns have been raised in recent years with respect to this technology. Autonomous vehicles can and do fail, such as in the prominent case of the fatal car crash of Chinese Tesla driver Gao Yaning (Boudette, 2016). However, with at least some failures being dependent on programming, publics tend to judge these failures more harshly than more spontaneous human failure. In particular, edge hypothetical cases have sparked discussion, such as those in which autonomous vehicles would potentially have to choose between two evils, such as *running over pedestrians or sacrificing themselves and their passengers to save the pedestrians (Nyholm & Smids, 2016). What makes such cases challenging and applicable to more comprehensive future implementation of autonomous agents is the necessity that programming these vehicles must include decision rules about what to do in such hypothetical situations beforehand, essentially binding manufacturers to decisions regarding who lives and who may get harmed. Manufacturers and regulators face the challenge of moderating a discussion of what kind of moral algorithms car owners are subjected to while not causing public outrage and delay adoption (Hanlon, 2016).

After considering robust action strategies for tacking algorithmic opacity, we will then use the final section of this paper to underpin robust pragmatic action with an additional normative layer to ensure not only robust, but also legitimate action.

PARTICIPATORY ARCHITECTURES FOR OPAQUE ALGORITHMS

Ferraro et al. (2015) proposed considering a structural dimension to tackling grand challenges; that is, architectures and forums of participation in which concerned actors may interact with each other. Due to the dynamic changes in complex algorithmic systems, the fostering of access to deliberation should ideally be supplemented

by such platforms that must allow for a sufficient continuity of debate (and not just for debate at selected time points) (Buhmann et al., 2019). Rigid certification processes, for instance, would not be able to do justice to the speed at which most complex algorithmic systems change. Recent suggestions for cooperative and procedural audits of algorithms (Mittelstadt, 2016; Sandvig et al., 2014a) directly addressed this aspect of continuity. The same aspect is also increasingly considered for public code repositories that use benchmark datasets to audit dynamic machine learning algorithms. This discussion indicates that deliberative forums for algorithmic accountability are likely to become an important area of contact and interaction between organisations and their environments. For the cases outlined above, consider that lay persons are widely shut off from any potential deliberation about CPSs. This is due to the fact that even to notice possible failure is widely impossible for non-specialists: "it remains highly unlikely that the failure will be evident to the data subject" (Mittelstadt, 2016, p. 4995). As a result, "deliberative audits" have been proposed in which service providers cooperate in processes that can create a record determining possible biases and help to e plain the ways in which people are profiled and why certain content is displayed to them (Mittelstadt, 2016; Sandvig et al., 2014a; Sandvig et al., 2014b). As these audits are cooperative and inclusive, they may serve as accessible platforms for deliberation about algorithmic accountability. Likewise, for autonomous vehicles, deliberations about moral decision rules are not open to the public—most likely because they involve the trade-off between mandatory ethics settings for the whole society and a driver's choice for his or her own personal ethics settings (Gogoll & Müller, 2017). A personal setting would most likely incur a prisoner's dilemma, as every driver would have a strong incentive to give priority to save him- or herself whenever possible. Therefore, deliberation is bound to specialist circles and the occasional media coverage that treats such dilemmas more as an interesting oddity than a matter of public discussion. All autonomous vehicles that are on the road are subject to constant review of regulators, such as the National Highway Traffic Safety Administration and open source systems exist with parts of their code available on repositories such as GitHub (comma.ai, 2017).

MULTIVOCAL INSCRIPTION FOR OPAQUE ALGORITHMS

For the instantiation of these discourses, or the inscription of differing viewpoints into material forms, we see an important obstacle to the realization of robust action strategies for tackling algorithmic opacity. Oftentimes the crucial information simply cannot be accessed, and even in cases where it can be accessed it may not be comprehensible in any sense that can serve as meaningful input for public debate. We established that potentially deliberative formats, such as audits or repositories, may in principle be hosted on open access platforms, but in practice still mostly give access to arguments of specialty audiences. This is a twofold problem: Algorithmic harm often arises from the way groups are classified or stigmatised. These groups are not only laypersons to algorithms; they are also often unaware that they are disadvantaged by them. This problem is currently only for the most severe instances, balanced by a public deliberation supported through platforms of the quality press and watchdog journalism: where concerns about algorithms become the object of broad public debate, the accountability discourse benefits, to some degree, from deliberation bolstered by quality media (see, e.g. Garber, 2016; Naughton, 2016; Smith, 2016 for the case of criminal justice algorithms). Specifically, such a high involvement of journalism points at both the access to deliberation as well as the inclusion of diverse arguments.

This, quite generally, highlights that holding algorithms accountable necessitates a responsive civil society that can feed diverse arguments into the debate (Kemper & Kolkman, 2019). Empowering agents, such as NGOs, regulators or civil society organisations, are essential for detecting and reviewing potential algorithmic failures and deliberating the intricate questions of accountability for multiple angles (Buhmann & Fieseler, 2021).

Consider that, for the cases mentioned, it has been stressed that information about the influence of CPSs which handle, for example, political information, has to be both accessible and comprehensible so that people can detect how their views may be externally shaped (Turilli & Floridi, 2009). As others have pointed out, it is insufficient to merely report on data features if the actual processes and logic behind the algorithms' decisions need to be understood (Burrell, 2016; Sandvig et al., 2014a). In many cases, as assessing the processes is difficult or practically impossible, information on CPSs is so far mostly gathered on the level of impact. It is conceivable that arguments from ordinary users can be included; for example, personalised prices on e-commerce platforms can be problematised when users share time-stamped prices (Mittelstadt, 2016). However, such arguments remain largely on the impact level of the algorithm and can hardly contribute to a debate on its opaque processes.

For autonomous vehicles, the different participants in such forums of deliberation can hardly operate on a level playing field. Even open source alternatives to navigational software, such as Comma.ai (a collection of software to enable autonomous vehicle navigation), are not fully open. Although the software is freely available, the code to core components is not. It lies in the safety-relevant nature of the technology that manufacturers are hesitant to make the software open to their owners, as they have strong incentives to alter the code to their benefit or game its algorithms. With the adoption of such vehicles, it is foreseeable that they are designed in a way that they only accept officially signed software and attempts to override or change them will become a felony. There are discussions in the engineering and computing communities on a more technical level, but on the consumer side, for the most part the former drivers are not included and their (however egoistic) concerns are unheard. The discourse is marked by rationality and socially optimal solutions, hence more egoistic concerns are dampened. New voices are brought to the discussion as the technology matures, with law enforcement agencies beginning to ponder possibilities of interception capabilities, for instance. There is also an emerging community of social media content creators that review and comment on the self-driving capabilities of their vehicle, with their voice to be expected to become stronger in the future.

DISTRIBUTED EXPERIMENTATION FOR OPAQUE ALGORITHMS

There is hope that ongoing public scrutiny will keep algorithmic organisations accountable. Increasing emphasis is put on policy that aligns with the common problems that many algorithms are currently struggling to address. However, distributed experimentation is a somewhat vague notion to guide organisational conduct (experiments performed to whose benefits?). Without moral guidelines, it is conceivable that action primarily benefits groups that have able proponents in the aforementioned forums, to the detriments of groups concerned along fault lines of, for instance, race, class, gender identity and sexual orientation, (dis)ability, language, or geographic location that may be primarily affected through algorithmic discrimination.

For the cases mentioned, responses to concerns about CPSs have, so far in almost every case, depended on the existence of regulatory bodies (Barocas & Selbst, 2015; Mittelstadt, 2016; Tutt, 2016). However, for the overall discourse to retain a high level of inclusiveness and engagement (and, thus, potentially legitimate outcomes), it seems necessary that such regulatory bodies are themselves charged with the creation of–and participation in–discursive platforms for algorithmic accountability. Otherwise, these bodies would be very powerful political actors that potentially distort discourses.

For autonomous vehicles, at the current moment, critical although unlikely edge cases are treated as moral precedents; that is, a socially beneficial default is assumed and not questioned. Discussions about the on-going development of the technology are hard to conduct, and manufacturers are both under the observation of authorities and must also safeguard their brand reputation when their software fails (as in the case of media coverage around Tesla autopilot failures). Thus, manufacturers are responsive to a degree to explain their technology, but not to the degree of independent core code review. Potential dangers such as the vehicles' vulnerabilities to hacking, furthermore, receive no response.

DISCUSSION: TOWARDS "PRINCIPLED ROBUST ACTION" FOR TACKLING ALGORITHMIC OPACITY

Quality of discursiveness and engagement

The preceding application of robust strategies to tackling the grand challenge of algorithmic opacity emphasises the limits of engagement and their according challenges to not only work towards novel solutions, but accountable and legitimate ones. The cases further show that the pragmatism-inspired calls for the inclusion of diverse and heterogeneous actors and the formation of diverse discourses and interpretations in robust action strategies rub up against the challenge of algorithmic opacity—as machine learning systems, which are different from older rule-based algorithms, are often not conducive to or designed with human understanding in mind (Edwards & Veale, 2017).

As Ferraro et al. (2015, p. 375) put it, "The key challenge for the focal actor is to prevent premature termination and to sustain engagement". The question then becomes, succinctly put: How can the inclusive engagement of diverse actors be sustained if neither developers, owners, nor users, can deliver viable accounts of processes and outcomes? This pronounces the importance of a closer look at the quality of discursiveness and engagement and thereby of the criteria on which such quality judgments can be based. However, so far such criteria that can serve as critical measures to assess the level and quality of engagement and discursiveness are not discussed explicitly in the literature on pragmatist approaches

to grand challenges. We argue that the grand challenge of algorithmic opacity points towards the necessity to extend the extant approach to robust action with a discussion of principles that can serve to evaluate organisational approaches of tackling grand challenges regarding their ability to foster not only novel, but legitimate, outcomes under the intricate conditions of algorithmic opacity.

We further argue that, so far, the concept of robust action has been discussed with an emphases on goal-driven action and the formation of power or success—be it in earlier work focussed more on individual actors (as in Eric Leifer's work on skill and chess strategy and Padgett and Ansell's (1993) work on Medicean achievement of political control) or in more recent work focussed on distributed action (as in Ferraro et al.'s 2015 and Etzion et al.'s 2017 approach to tackling grand challenges). While Ferraro et al. (2015, p. 10) maintain that their approach proposes strategies that "can be harnessed for positive distributed outcomes, rather than for individual gain", the focus remains mainly with experimentalism and general enablers of a robust progress for innovation rather than explicating specifics of such "positive outcomes" and their emergence vis-a-vis societal needs and expectations. While the pragmatist approach calls for engagement and discursiveness to generate "small wins", the approach does not go into detail regarding the ways in which we can decide whether we can justifiably speak of "discursiveness" or "engagement", and, ultimately, know under what conditions one could justifiably speak of a "win" for those partaking in the distributed experimentation. In short: what actually bolsters 'true' engagement and discursiveness? Put a different way: So far, participation and multivocality are intended to enable a form of experimental "free play" that works towards novel solutions and sustained engagement. But what makes these solutions legitimate and how can legitimation be approached specifically in the face of the challenge of algorithmic opacity? To address these questions, in the following two sections, we foreground a recent discussion on the legitimation of algorithmic systems and the organizations that develop and employ them (Buhmann et al., 2020) as a potential amendment to the robust action approach.

STRATEGIES OF ORGANISATIONAL LEGITIMATION

Following Scherer, Palazzo, and Seidl (2013), we argue that organisations have three fundamental strategic options to foster legitimacy through engagement: they can a) strategically manipulate expectations, b) adapt and conform to extant expectations in their environment, or c) engage public debate and reasoning over what should be expected. The manipulative approach describes the active attempt to shape and influence external expectations, such as through lobbying, public relations campaigns and o her strategic communication instruments. This approach is guided not by adherence to external demands or institutional rights to information, but rather by the solicitation of stakeholder views in a reputational contest for the sake of reputation, thus leading to 'soft accountability' (Owen et al., 2000; Swift, 2001).

The adaptive approach describes isomorphic behaviour aimed to conform with extant expectations through meeting the demands of powerful stakeholders or complying with established standards (e.g. leading to practices of reporting or performance review). Through a reputational lens, this emphasises an outside-in approach beyond mere influence in which stakeholder partnerships facilitate organisational learning and the adjustment of main reputation drivers (Romenti, 2010). Of course, for this approach to work, external expectations have to be rather clear-cut and stable, which, in the case of self-learning machines that are based on algorithms, can hardly be assumed.

Finally, the moral approach builds on open discourse between the organisation and its stakeholders and free exchange of arguments that can lead to common outcomes in terms of what should be expected. As such, a moral approach helps to facilitate legitimate outcomes under conditions of unclear external demands (Mingers & Walsham, 2010) where knowledge about the workings and ramifications of algorithms do not exclusively reside within the organisation, but must emerge from an open deliberation with actors in the organisation's environment who are affected by it (Lubit, 2001).

For organisations, these three fundamental strategic options constitute parallel approaches, rather than mutually exclusive strategies. Depending on the particular challenge at hand, they can be enacted simultaneously (Scherer et al., 2013). While robust action strategies seem to fit with the more interaction-based approach that works towards 'moral legitimacy' by including diverse stakeholder perspectives and enabling discursiveness, there is so far no discussion regarding criteria that would allow to assess the degree to which applied robust action strategies actually enable legitimate solutions as an outcome of engagement.

TOWARDS PRINCIPLES FOR LEGITIMATE NOVELTY

When strategies for participatory architectures, multivocal inscription and distributed experimentation are faced with the challenge of algorithmic opacity, it highlights the importance to extend the discussion towards the quality of engagement and explicate the degree to which such engagement can serve as a communication process through which a continuous and tentative assessment of the development, workings, and consequences of algorithms can be achieved over time. Following similar applications (cf. Palazzo & Scherer, 2006; Seele & Lock, 2015; Gilbert & Rasche, 2007; Rasche & Esser, 2006), we have recently suggested (Buhmann et al. 2020) to draw on a discourse-ethical approach to derive communicative principles that allow the further assessment of the quality of engagement in approached aimed at algorithmic accountability. In their most widely used form, discourse-ethical approaches draw on Habermas' (1999) work on discourse about competing validity claims, in which participants consider each other's arguments, give reasons for their position, and are ultimately willing to reassess and, if necessary, revise their original position. Such a discourse leads to a deeper understanding of the problems, positions and concerns of the various actors, as well as a greater mutual acceptance of all parties involved and the common (ideally consensual) decisions. However, the possibility of such positive outcomes hinges on the adherence to normative principles when debating the acceptance or rejection of particular validity claims, such as the principle of open and equal access to forums of discussion, the availability and transparency of information, and equal opportunities for all to introduce arguments into the debate.

These communicative principles are to ensure that discourses are un-corrupted by power differences or strategic motivations (see, e.g., Niemi, 2008, for a concise summary of the approach). As outlined in Buhmann at al. (2020) based on Nanz and Steffek's work (2005) algorithmic accountability can be addressed through communicative principles of participation, comprehension, multivocality, and responsiveness. The principle of participation asks that intricate issues around algorithmic accountability be discussed in an open forum in which all subjects with the competence to speak and act (specifically, all those who potentially suffer negative effects of the processes and decisions of algorithmic systems) are allowed to take part in the debate. This debate should aim to spotlight potential issues, facilitate argumentation, and lead to broadly acceptable decisions. Second, the principle of comprehension asks that all those who participate in the deliberative process have full information about the issues at stake, the various suggestions for their solution and the ramifications of these suggestions. Third, the principle of multivocality asks for an open and dialogic inclusion of all arguments to enable rational discourse and deliberation (i.e., participants aim to see an issue from all relevant points of view and remain open to revising their own point of view based on the most convincing arguments). Fourth, while participation and comprehension are preconditions for a process of deliberation to take place and the inclusion of all arguments is the main precondition of the rationality of that process, the principle of responsiveness asks for the different concerns and suggestions regarding algorithmic systems that are put forth by various stakeholders to be adequately taken up in the actual recommendations or decisions that emerge as the result of the discourse (see figure 1).

TOWARDS LEGITIMATE NOVELTY THROUGH PRINCIPLED COMMUNICATIVE ENGAGEMENT

Based on the above discussion, such a discourse-ethical approach based on communicative principles seems fitting on two levels. On the one hand, in light of the challenges of algorithmic opacity, developers and proprietors of machine learning systems need to be prepared to participate in a discursive process together with their stakeholders in order to work towards the accountability of algorithms. Discourse principles place a strong emphasis on involving those affected by decisions. Stakeholders need to be an active part of detecting and assessing the potential shortcomings of algorithms, as particular developers and applicants of algorithms do not necessarily hold a privileged position in assessing these issues. Accordingly, discourse principles are vital to addressing accountability in the context of highly fluid and constantly evolving information systems Buhmann et al., 2020; cf. also Mingers & Walsham, 2010).

On the other hand, in light of the application of a robust action approach to tackling grand challenges, discourse- ethical principles for rational communication serve as a suitable addition as they are developed, not only according to the same general pragmatist convictions of the value and vulnerability of intersubjective, reciprocal, egalitar an communication, but also as a specific theoretical extension of pragmatist conceptions of rationality (Bernstein, 1992). As discussed famously by Bernstein

(1992), the idea of communicative rationality was first developed by Peirce in his self-corrective community of critical inquirers; it continues in Dewey's ideas of democracy and in Mead's discussion of the institutionalisation of democratic forms of life. Rationality, as an essentially dialogical and communicative element, has been a core theme both for the pragmatic tradition and for neo-pragmatists (cf. Putnam, 1981; Joas, 1993; Habermas, 2002) and, as we argue, can thus serve as a normative extension to current work on robust action that not only acknowledges the necessity to "ensure legitimacy", but points specifically to related conceptual work (e.g. by Mena and Palazzo, 2012) that applies such communicative principles (Ferraro et al., 2015, p. 375).

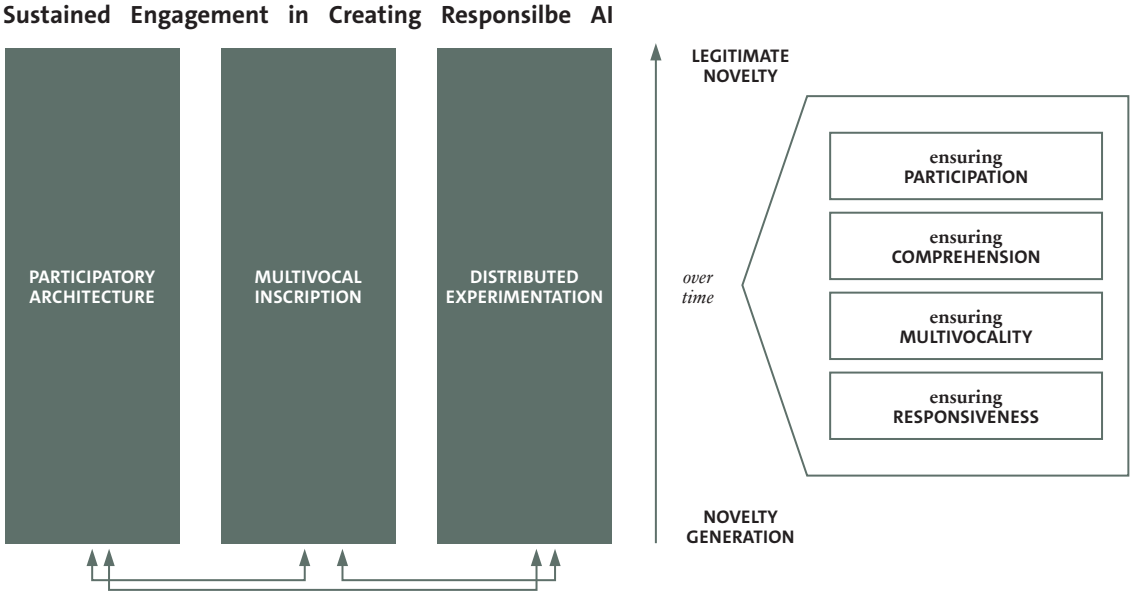

Figure 1: Towards Legitimate Novelty through Principled Communicative Engagement

CONCLUSION:

We argued that robust action approaches are conducive to tackling the grand challenge of algorithmic opacity through participation and experimentation. However, in light of the specific technical and procedural challenges to algorithmic accountability, these strategies should be amended with a set of communicative principles that allow us to assess the quality of engagement and discursiveness from a specific normative perspective. As Ferraro et al. (2015, p. 371), following Ansell's (2011) work on pragmatism and evolutionary learning, have stressed: problem-solving, reflexivity, and deliberation "need to work together in a recursive fashion for evolutionary learning to occur". We add to this the notion that the potential of deliberation hinges on the ability of the engaged actors to foster a rational debate by adhering to communicative principles. While the extant conceptual work on robust action has emphasised iterative action that increases engagement, discursiveness to sustain different interpretations, and "rules of engagement that allow diverse and heterogeneous actors to interact constructively over prolonged timespans" (Ferraro et al., 2015, p. 374), so far no such rules of engagement and discursiveness have been further developed within the pragmatist approach to tackle grand challenges. At the point the approach has remained rather 'open' regarding specifics of engagement and discursiveness, as the main focus was to "mobilize heterogeneous actors and generate novel solutions" (Ferraro et al. 2015, p. 366). To a good degree, it relies on an intuitive adherence to a 'liberal culture' and largely abstains from a more specific debate about principles that could ensure communicative rationality in the process. Rather than going into principles to further support notions of engagement and discursiveness, extant work turns, more or less intuitively, to the contextual, to extant habits and practices. Processes of engagement and discourse tend to be more loosely seen as inherently ideal rather than amending them with some universal defence of specific ideals. To some, such an 'anti-foundationalist' approach may be what pragmatism is all about: It has been stressed that the Habermasian non-teleological conception of communicative action starts with too-strong assumptions of the rationality of action (Joas, 1993) and retains too much "Kantian transcendentalism" (Margolis, 2002; Rockmore, 2002) to be "truly pragmatist". However, with Bernstein and Habermas' own characterisations, we maintain that the principles and universality underlying the work on communicative action remains entirely procedural and, thus, is in line with a pragmatist notion of fallibility. As such, communicative principles serve as a valuable addition to current pragmatist approaches for tackling grand challenges and substantiating a specific point of critique from which actual engagement and discursiveness in robust action strategies can be assessed.

REFERENCES:

Ananny, M. (2016). Toward n ethics of algorithms: Convening, observation, probability, and timeliness. Science, Technology, Human Values, 41(1), 93-117.

Ananny, M., & Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, (3)2, 1-17.

Ansell, C. (2011). Pragmatist democracy: Evolutionary learning as public philosophy. New York, NY: Oxford University Press.

Bagloee, S. A., Tavana, M., Asadi, M., & Oliver, T. (2016). Autonomous vehicles: challenges, opportunities, and future implications for transportation policies. Journal of Modern Transportation, 24(4), 284-303.

Barnet, B.A. (2009). Idiomedia: The rise of personalized, aggregated content. Continuum 23(1), 93–99.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104, 671-732.

Beck, M. (2016). Can a Death-Predicting Algorithm Improve Care? Wall Street Journal, 2. December 2016.

Beer, D. (2009). Power through the Algorithm? Participatory Web Cultures and the Technological Unconscious. New Media & Society, 11(6), 985-1002.

Beer, D. (2013). Popular culture and new media: The politics of circulation. Basingstoke: Palgrave Macmillan.

Bernstein, R. J. (1992). The new constellation: The ethical-political horizons of modernity/postmodernity. MIT press.

Boudette, N. (2016): Autopilot Cited in Death of Chinese Tesla Driver. New York Times. https://www.nytimes.com/2016/09/15/business/fatal-tesla-crash-in-china-involved-autopilot-government-tv-says.html

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. New Media & Society, 14(7), 1164-1180.

Buhmann, A., & Fieseler, C. (2021). Towards a deliberative framework for responsible innovation in artificial intelligence. Technology in Society, 64, 101475.

Buhmann, A., Paßmann, J., & Fieseler, C. (2020). Managing Algorithmic Accountability: Balancing Reputational Concerns, Engagement Strategies, and the Potential of Rational Discourse. Journal of business ethics, 163(2), 265-280.

Burrell, J. (2016): How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 1-17.

Comma.ai (2017). Comma.ai software repository. Available at: https://github.com/comma-ai (accessed December 14th, 2017).

Dentoni, D., Bitzer, V., & Schouten, G. (2018). Harnessing wicked problems in multi-stakeholder partnerships. Journal of Business Ethics, 1-24.

Edwards, L. (2016). The role of public relations in deliberative systems. Journal of communication, 66(1), 60-81.

Edwards, L., & Veale, M. (2017). Slave to the algorithm?. Why a 'Right to Explanation' is Probably Not the Remedy You Are Looking For. 16 Duke Law & Technology Review 18 (2017). Available at SSRN: https://ssrn.com/abstract=2972855 or http://dx.doi.org/10.2139/ssrn.2972855

Etzion, D., Gehman, J., Ferraro, F., & Avidan, M. (2017). Unleashing sustainability transformations through robust action. Journal of Cleaner Production, 140, 167-178.

Ferraro, F., Etzion, D., & Gehman, J. (2015). Tackling grand challenges pragmatically: Robust action revisited. Organization Studies, 36(3), 363-390.

Garber, M. (2016). When Algorithms Take the Stand, The Atlantic. June 30, 2016.

George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2016). Understanding and tackling societal grand challenges through management research. Academy of Management Journal, 59(6), 1880.

Gilbert, D. U., & Rasche, A. (2007). Discourse Ethics and Social Accountability: The Ethics of SA 8000. Business Ethics Quarterly, 17(2), 187-216.

Gilbert, D. U., & Rasche, A. (2007). Discourse Ethics and Social Accountability: The Ethics of SA 8000. Business Ethics Quarterly, 17(2), 187-216.

Gillespie, T. (2014). The Relevance of Algorithms. In T. Gillespie, P.J. Boczkowski & K.A. Foot (Eds.), Media Technologies. Essays on Communication, Materiality, and Society, Cambrdige/MA: MIT Press, pp. 167-194.

Glenn, T., & Monteith, S. (2014). Privacy in the digital world: medical and health data outside of HIPAA protections. Current Psychiatry Reports, 16(11), 494, 1-11.

Gogoll, J., & Müller, J. F. (2017). Autonomous cars: in favor of a mandatory ethics setting. Science and Engineering Ethics, 23(3), 681-700. Goldman, E. (2006). Search engine bias and the demise of search engine utopianism. Yale Journal of Law & Technology, 8(1), 6–8.

Gray, R. (2002). The social accounting project and Accounting Organizations and Society Privileging engagement, imaginings, new accountings and pragmatism over critique? Accounting, Organizations and Society, 27(7), 687-708. doi:https://doi.org/10.1016/S0361-3682(00)00003-9

Greenwood, M. (2007). Stakeholder Engagement: Beyond the Myth of Corporate Responsibility. Journal of Business Ethics, 74(4), 315-327.

Greenwood, R., Raynard, M., Kodeih, F., Micelotta, E. R., & Lounsbury, M. (2011). Institutional Complexity and Organizational Responses. Academy of Management Annals, 5(1), 317-371.

Grodal, S., & O'Mahony, S. (2017). How does a grand challenge become displaced? Explaining the duality of field mobilization. Academy of Management Journal, 60(5), 1801-1827.

Habermas, J (2002): "Postscript: some concluding remarks". In M. Aboulafia, M Bookman, & C. Kemp (Eds.), Habermas and pragmatism (pp. 223-233). London, UK: Routledge.

Habermas, J. (1999). Moral Consciousness and Communicative Action (trans. Christian Lenhardt, Shierry Weber Nicholsen). Cambridge, Mass.: MIT Press.

Hanlon, M. L. (2016). Self-Driving Cars: Autonomous Technology that Needs Designated Duty Passenger. Barry L. Rev., 22, 1.

Hildebrandt, M. (2008). Profiling and the rule of law. Identity in the Information Society, 1(1), 55-70.

Joas, H. (1993). Pragmatism and social theory. Chicago and London: University of Chicago Press.

Kehl, D., Guo, P., & Kessler, S. (2017). Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Berkman Klein Center for Internet & Society, Harvard Law School. Available at: https://cyber.harvard.edu/ publications/2017/07/Algorithms (accessed September 5th, 2018)

Kemper, J., & Kolkman, D. 2019. Transparent to whom? No algorithmic accountability without a critical audience. Information, Communication & Society, 22(14): 2081–2096.

Kim, H., Giacomin, J., & Macredie, R. (2014). A qualitative study of stakeholders' perspectives on the social network service environment. International Journal of Human-Computer Interaction, 30(12), 965-976.

Leese, M. (2014). The new profiling: Algorithms, black boxes, and the failure of anti - discriminatory safeguards in the European Union. Security Dialogue , 45(5), 494 – 511.

Lubit, R. (2001). The keys to sustainable competitive advantage: Tacit knowledge and knowledge management. Organizational Dynamics, 29(3), 164-178.

Margolis, J. (2002). Vicissitudes of transcendental reason. In M. Aboulafia, M Bookman, & C. Kemp (Eds.), Habermas and pragmatism (pp. 31-46). London, UK: Routledge.

Martí, I. (2018). Transformational business models, grand challenges, and social impact. Journal of Business Ethics, Online First, pp. 1-12. doi.org/10.1007/s10551-018-3824-3 Martin, K. (2018). Ethical Implications and Accountability of Algorithms. Journal of Business Ethics, Online First, pp. 1-16, doi:10.1007/s10551-018-3921-3.

Mena, S., & Palazzo, G. (2012). Input and output legitimacy

of multi-stakeholder initiatives. Business Ethics Quarterly, 22, 527–556.

Mingers, J., & Walsham, G. (2010). Toward ethical information systems: the contribution of discourse ethics. MIS Quarterly, 34(4), 833-85

Mittelstadt, B. (2016). Auditing for Transparency in Content Personalization Systems. International Journal of Communication, 10, 4991-5002.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1-21.

Nanz, P., & Steffek, J. (2005). Assessing the Democratic Quality of Deliberation in International Governance: Criteria and Research Strategies. Acta Politica 40, 368-383.

Naughton, J. (2016). Opinion, Even Algorithms Are Biased Against Black Men. The Guardian. June 26, 2016.

Niemi, J. I. (2008). The foundations of Jürgen Habermas's discourse ethics. The Journal of Value Inquiry, 42(2), 255-268.

Nyholm, S., & Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: an applied trolley problem?. Ethical Theory and Moral Practice, 19(5), 1275-1289.

Owen, D. L., Swift, T. A., Humphrey, C., & Bowerman, M. (2000). The new social audits: accountability, managerial capture or the agenda of social champions? European Accounting Review, 9(1), 81-98.

Padgett, J. F., & Ansell, C. K. (1993). Robust action and the rise of the Medici, 1400–1434. American Journal of Sociology, 98, 1259–1319.

Palazzo, G., & Scherer, A. G. (2006). Corporate legitimacy as deliberation: A communicative framework. Journal of Business Ethics, 66(1), 71-88.

Pasquale, F. (2015). The Black Box Society: The Secret Algorithms that Control Money and Information. Cambridge, MA: Harvard University Press.

Putnam, H. (1981). Reason, truth, and history. Cambridge: Cambridge University Press.

Rasche, A., & Esser, D. (2006). From Stakeholder Management to Stakeholder Accountability Applying Habermasian Discourse Ethics to Accountability Research. Journal of Business Ethics, 65(3), 251–267.

Rockmore, T. (2002). The epistemological promise of pragmatism. In M. Aboulafia, M Bookman, & C. Kemp (Eds.), Habermas and pragmatism (pp. 47-64). London, UK: Routledge.

Romenti, S. (2010). Reputation and stakeholder engagement: an Italian case study. Journal of Communication Management, 14(4), 306-318.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014a). An Algorithm Audit. In: S.P. Gangadharan (ed.), Data and Discrimination: Collected Essays (pp. 6-10). Washington, DC: New America Foundation.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014b). Auditing algorithms: Research methods for detecting discrimination on internet platforms. Paper presented to "Data and Discrimination: Converting Critical Concerns into Productive Inquiry," a preconference at the 64th Annual Meeting of the International Communication Association. May 22, 2014; Seattle, WA, USA.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2016). When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software. International Journal of Communication, 10, 4972-4990.

Scherer, A. G., Palazzo, G. and Seidl, D. (2013). 'Managing legitimacy in complex and heterogeneous environments: sustainable development in a globalized world.' Journal of Management Studies, 50, 259–84.

Seele, P. & Lock, I. (2015). Instrumental and/or Deliberative? A Typology of CSR Communication Tools. Journal of Business Ethics, 131(2), 401-414.

Smith, M. (2016). In Wisconsin, a Backlash Against Using Data to Foretell Defendants' Futures. NY Times. June 22, 2016.

Stark, M., & Fins, J. J. (2013). What's Not Being Shared in Shared Decision- Making?. Hastings Center Report, 43(4), 13-16.

Steenbergen, M. R., Bachtiger, A., Sporndli, M., & Steiner, J. (2003). Measuring Political Deliberation: A Discourse Quality Index. Comparative European Politics, 1(1), 21-48.

Swift, T. (2001). Trust, reputation and corporate accountability to stakeholders. Business Ethics, a European Review, 10(1), 16-26.

Turilli, M., & Floridi, L. (2009). The ethics of information transparency. Ethics and Information Technology, 11(2), 105–112.

Tutt, A. (2016). An FDA for algorithms. Social Science Research Network. Available at http://papers.ssrn.com/abstract=2747994 (accessed October 6th, 2018).

Van Buren, H. J. (2001). If Fairness is the Problem, Is Consent the Solution? Integrating ISCT and Stakeholder Theory. Business Ethics Quarterly, 11(3), 481-499.

Woolley, S. (2016). Automating power: Social bot interference in global politics. First Monday, 21(4).

Zarsky, T. (2016). The trouble with algorithmic decisions an analytic road map to examine efficiency and fairness in automated and opaque decision making. Science, Technology & Human Values, 41(1), 118–132.