AI & SOCIETY (2024) 39:195–205
https://doi.org/10.1007/s00146-022-01395-1

ORIGINAL PAPER

Challenges in enabling user control over algorithm-based services

Pascal D. König[1]

Received: 6 August 2021 / Accepted: 31 January 2022 / Published online: 24 February 2022

Abstract
Algorithmic systems that provide services to people by supporting or replacing human decision-making promise greater convenience in various areas. The opacity of these applications, however, means that it is not clear how much they truly serve their users. A promising way to address the issue of possible undesired biases consists in giving users control by letting them configure a system and aligning its performance with users' own preferences. However, as the present paper argues, this form of control over an algorithmic system demands an algorithmic literacy that also entails a certain way of making oneself knowable: users must interrogate their own dispositions and see how these can be formalized such that they can be translated into the algorithmic system. This may, however, extend already existing practices through which people are monitored and probed and means that exerting such control requires users to direct a computational mode of thinking at themselves.

Keywords Algorithmic decision-making · Algorithm bias · Algorithmic literacy · Surveillance

1 Introduction

Information societies see an increasing adoption of algorithmic decision-making (ADM) systems that promise to contribute to better-informed decisions and to greater convenience. ADM systems are employed in a broad range of domains and with variable relationships between the algorithms and the data subjects who are classified or profiled in some form. Where algorithms produce risk assessments in credit lending or in criminal justice, for instance, individuals are objects of decisions but not the primary users of these systems who receive some form of service. In other contexts, individuals in their role as consumers or citizens are the primary users themselves. This is the case, e.g., where they rely on AI to filter online content for them, match people in online dating platforms, make medical diagnoses or predictions, or serve as personal assistants. The present paper is concerned with this second kind of applications which provide a service to consumers or citizens as primary users.

The focus will lie on recommender systems as a widespread form of ADM systems that many people encounter in their daily lives. While these provide a service, e.g., by helping users navigate online environments, they are, however, commonly opaque and users have no insight into which goals an algorithmic system optimizes exactly (Pasquale 2015; Binns 2017; Ananny and Crawford 2018). It thus remains unclear whether and how exactly they serve those users. This issue briefly gained traction in public debates in 2021, when the content filtering algorithms on Facebook and Instagram came under scrutiny (Metz 2021): While these enabled people to connect with others and to navigate online content, the algorithmic filters maximized attention and involvement through prioritizing content which, among other things, evoked negative emotions. The problem is thus that behind users' backs, algorithms may operate in ways that produce undesirable side-effects and biases.

One should note that all ADM systems will be biased in the general sense that their decision models are based on learned relationships and that they necessarily incorporate certain assumptions and values (Mittelstadt et al. 2016). The kind of bias that is of interest in the following refers to cases in which ADM systems explicitly or implicitly realize goals that diverge from the interests of those to whom an ADM system is supposed to provide a service. Research on accountable algorithms has devised various ways in which ADM systems can be made more transparent and tested to make sure that they do not perform in an undesirable fashion (e.g. Diakopoulos 2014; Krafft et al. 2020; Kroll et al. 2017; Lepri et al. 2018; Felzmann

✉ Pascal D. König
pascal.koenig@sowi.uni-kl.de

[1] Department of Social Sciences, TU Kaiserslautern, Building 57, PO-Box 3049, 67653 Kaiserslautern, Germany

et al. 2020). Through realizing features of explainable AI, users of algorithm-based services can also be enabled to receive information on how a certain recommendation or decision has been arrived at and under which conditions they would have turned out differently (Holzinger et al. 2018; Samek and Müller 2019). Going beyond these ways of making ADM systems transparent and accountable, there is also the idea of more directly letting the users of algorithm-based services exert control over the performance of algorithmic systems: via design features that allow users to configure algorithm-based services to their liking and thus to align their performance with users' own preferences (e.g. Harper et al. 2015; van Drunen et al. 2019; Harambam et al. 2019). Doing so promises to avoid the problem of undesirable biases in algorithmic systems. Enabling control over algorithms in this sense would thus appear as a straightforward solution for levelling information and power asymmetries in the—increasingly common—instances in which citizens and consumers rely on algorithm-based services. It would mean that they could directly tackle and correct the forms of bias as defined above and ensure that ADM systems are aligned with their own interests.

However, as the present paper argues, this approach comes at the price of also having to exert a certain kind of discipline over the self. In a nutshell, configuring an algorithmic system by aligning its performance with one's own goals and values demands a literacy of such systems that also entails a specific way of making oneself knowable: Individuals must interrogate themselves as they learn to perceive their own dispositions in such a way that they can be translated into an ADM system. There is thus a dialectic involved in user control over algorithms. It implies that the algorithmic literacy on which such control is based is not merely a neutral knowledge that individuals can use for their purposes. Rather, it amounts to a way of perceiving oneself through metrics that render one's goals and values formalizable. While this could be empowering under certain conditions, it generally means that the price for control is being subjected to a specific form of self-knowledge. This dialectic will be discussed focusing on the example of recommender systems and specifically by taking up the example of content filtering algorithms.

Through developing this argument, the present paper adds an important facet to existing work on algorithmic literacy as an ability needed to competently deal with ADM systems (D'Ignazio and Bhargava 2015; Baker 2017; Klawitter and Hargittai 2018; Zhu et al. 2018; Lloyd 2019; Cotter and Reisdorf 2020; Hargittai et al. 2020; Sander 2020; Bakke 2020). The discussion below shows in what sense the literacy needed to exercise control over algorithms can end up extending an already prevalent tendency of measuring and quantifying the self, namely as user control over algorithms demands of individuals that they interrogate their own dispositions and personality under a perspective that subjects them to a computational mode of thinking.

2 Agency problems in the use of algorithmic systems

2.1 Serving Whom?

The potential of ADM systems to augment or replace human decision-making stems from an ability to solve analytical tasks based on complex decision rules and through processing large amounts of data (Mittelstadt et al. 2016, p. 3). This ability is rooted in machine learning methods through which algorithmic systems acquire—and update—an optimal decision model from patterns registered in input data. A trained decision model for dealing, e.g., with a classification or sorting task may then form the basis of a concrete service. As such, algorithmic systems may support people with the filtering and selection of online news, product choices, health-related recommendations etc. Users of such services can be presumed to have an interest in these algorithmic systems performing well. However, the question is what performing well means in the first place.

ADM systems are never neutral technological implementations. They necessarily come with assumptions and objectives embedded into them (Hildebrandt 2016; Yeung 2017a). These define what counts as an optimal solution for a given task and guide how an algorithmic system 'learns' decision rules from input data and ultimately produces decisions. ADM systems are therefore always biased in the general sense that they prioritize and realize certain objectives rather than others, which can be a conscious or an unintended part of the ADM system design (Barocas and Selbst 2016; Mittelstadt et al. 2016; Lepri et al. 2018). At the same time, ADM systems commonly remain highly opaque, with central parameters, optimized goals, and performance metrics remaining hidden (Pasquale 2015; Binns 2017; Ananny and Crawford 2018). Users of algorithm-based services, therefore, commonly cannot evaluate how much an ADM system truly serves their interests or might instead show a bias that goes against their own goals.

The prevailing practice is that users can expect increased convenience from algorithm-based services while not being bothered with questions regarding which goals and values a system actually realizes (Zuboff 2019). Users are supposed to blindly rely on these systems and can only trust that these systems operate in their best interest (Lepri et al. 2018; van Drunen et al. 2019, pp. 1, 5). This may mean, however, that individuals are unwittingly steered toward choices in line with predefined objectives and according to desires that cannot be understood as authentically their own (Yeung 2017b;

Lanzing 2018). A social network news and content filtering algorithm, for example, might be designed to give greater weight to polarizing content and negativity to generate more attention and involvement. For this to happen, the algorithm does not have to be directly designed in a way such that it prioritizes negative content. Rather, this can be an unintended consequence if an algorithm merely maximizes attention and engagement, which are themselves higher as a result of negative content.

In any case, as Helberger et al. (2018, p. 9) have remarked with regard to algorithmic content filters in social media: '[T]hat large, extremely opaque, and primarily profit-driven data companies should determine what (and what does not) constitutes a healthy (i.e., diverse) media diet is clearly problematic'. Similar problems arise with more specialized applications that produce recommendations for their users: Baker (2017) has noted with regard to algorithms created for navigating legal databases that different applications may vary considerably in the relevant results delivered after a query. A blind reliance on a specific system by a user from the legal profession will thus mean that she 'has just allowed the algorithm to have a significant role in selecting the cases the algorithm deems should advance the law' (Baker 2017, p. 572). In this case, too, the user of an algorithmic system must trust that the algorithm is based on goals and notions of relevance that are in line with the user's own—at the risk that there is an undesirable bias at work.

In sum, while ADM systems may seemingly serve to better give individuals what they want, this may take place under conditions over which these individuals have neither knowledge nor control. It may often remain hidden to them that an ADM system shows an undesirable bias as it realizes objectives and values that deviate from users' preferences.

2.2 Beyond accountability

In light of the challenges described above, a growing literature is dealing with ways of achieving ethically acceptable and accountable algorithms (e.g. Ananny and Crawford 2018; Reisman et al. 2018; Lepri et al. 2018). Various technical means have been proposed for realizing transparency and accountability through ascertaining whether certain standards (e.g. regarding a specific fairness conception) and criteria are realized by an ADM system. This may occur through providing relevant information about the design of the system or by testing and auditing ADM systems to detect possible undesirable biases or other flaws (Diakopoulos 2014; Kroll et al. 2017; Wachter et al. 2017; e.g. Bryson and Theodorou 2019).

These instruments will in large part have to be used by third parties, such as regulators or civil society actors (Saurwein et al. 2015). The literature, however, also points to ways in which individuals as users can be enabled to influence how an ADM system operates. They point to the possibility of introducing forms of direct user control as users themselves are enabled to configure which goals an ADM system realizes and how it trades off certain objectives or performance criteria against each other. This can be achieved through design features which allow users not only to monitor and assess the performance of algorithm-based services but also to configure them by changing relevant parameters. It has been noted that suitable interfaces and dashboards can help users in this regard, through showing them how an algorithmic system performs (van Drunen et al. 2019, p. 14; Matheus et al. 2020; Yu et al. 2020), and the functionality of a recommender system could let users experiment with, tune and set certain parameters (such as the importance of recency and popularity of content) that affect their recommendation results (Harper et al. 2015). In the context of social media and algorithmic content selection, it has been argued that the possibility of altering performance-relevant parameters of the ADM system would allow users to customize how decisions are made for them (Helberger et al. 2018). For instance, users could be provided information—expressed in suitable metrics—about the diversity of their content exposure, and be enabled to modify the filtering logic and criteria (Helberger et al. 2018, p. 10; see also Harambam et al. 2019).

Indeed, recent empirical work on people trying to exert control over algorithms, which is predominantly set in the context of social media content selection, suggests that at least some users seek control over algorithmic processes (e.g. Burrell et al. 2019), but that users also often find themselves at a loss, are unaware of the content selection logic (Hsu et al. 2020), and that even placebo control instruments may increase user satisfaction (Vaccaro et al. 2018). Research also points to an overall low degree of user awareness regarding algorithms operating in online platforms or other applications and to widespread misconceptions about algorithms—which are, however, unevenly distributed and linked to various sociodemographic characteristics (Eslami et al. 2016; Gran et al. 2021; Zarouali et al. 2021). There are thus signs of a new digital divide that indicates a need for fostering algorithmic literacy throughout society. Further, users of online content filters commonly feel irritated and regularly express concern about a lack of transparency and possible biases while perceiving algorithms as inescapable (Ytre-Arne and Moe 2021). There is thus considerable scope for better enabling individuals to effectively exert control over algorithms in online environments through configuring them according to their liking. This would similarly work with other algorithm-based services and in other domains, such as voice-controlled personal assistants.

In any case, users would face the task of finding a configuration of the application that best accommodates their preferences as they attempt to translate their own goals and values into the operating of the system. Such direct control over the behavior of ADM systems goes beyond establishing transparency and accountability and allows individuals to avoid undesirable biases through actively aligning the functioning of ADM systems with their own dispositions. To do so, however, they must be able to understand how the performance and outputs produced by an ADM system relate to their own goals and values. To address and minimize the problem of possible undesirable biases, they must understand how any conceivable undesired biases and corresponding configuration of the ADM system relate to their own preferences. In this sense, control over algorithmic systems demands a certain algorithmic literacy.

2.3 The role of algorithmic literacy in control over algorithms

Already without possibilities to directly influence the performance of algorithmic systems, a general understanding of these systems is crucial. For effective accountability of algorithms, more is needed than merely transparency about the system, e.g. regarding its basic logic and possibly even the decision model and weights of decision criteria (Malgieri 2019). There also has to be intelligibility without which transparency or the explanation of decision-making are useless (Edwards and Veale 2017; Malgieri and Comandé 2017). In other words, data subjects need to be able to understand what any provided information about the functioning of an ADM system means.

To thoroughly understand algorithm-based services that individuals engage with, it therefore seems that they require a certain algorithmic literacy. Such a literacy is different from merely a computer literacy as the ability to make use of and 'function independently with a computer' (Robinson 2009, p. 128) and from a data literacy as an ability to read, interpret, manage, analyze, and argue with data (Calzada-Prado and Marzal 2013, pp. 124–125). Rather, algorithmic literacy has been treated as an empowering ability to critically examine how one interacts with algorithmic systems, such as search engines, and how these impact on one's agency (Sander 2020; Bakke 2020). This may entail a knowledge of how assumptions and biases are inherent in their construction (Lloyd 2019, pp. 1482–1483) and a basic understanding of how algorithms produce outputs based on data inputs (D'Ignazio and Bhargava 2015, p. 3; Rainie and Anderson 2017, p. 15; Cotter and Reisdorf 2020, p. 747).

Such algorithmic literacy equips individuals with a reflective stance toward algorithmic systems, enabling them to see how the outputs obtained from algorithmic systems may express the goals, interests, and assumptions of others. However, the empowering effect of such a critical understanding of algorithms is limited for two reasons. *First*, literacy does not mean that people can also use this understanding to exert actual control over the operations of an algorithmic system (e.g. through a design that allows for configuring its performance)—literacy, just like transparency, can be "disconnected from power" (Ananny and Crawford 2018, p. 6). Even if this challenge is overcome through giving users more control via individual influence over its configuration and behavior there remains a *second* obstacle: If users are given this kind of power, they need more than a general and critical understanding of algorithms. Directly achieving control over algorithm-based services as described above would mean aligning its operations with one's own preferences and values and thus requires a translation of one's own dispositions into the functioning and the performance of an algorithmic system. The kind of algorithmic literacy required for this kind of control thus comprises *the ability to understand how to modify performance-relevant parameters of an ADM system in such a way that it best realizes one's own goals and values.*

Hence, to minimize undesired biases, individuals must be able to find answers to questions such as: What would be alternative ways in which the system could operate? How would it have to be configured, which weights do certain criteria have to be given such that one's values and preferences are better reflected in the system's performance? In sum, what is a "good" performance to me expressed in relevant parameters that guide the ADM system? It is this kind of literacy that enables individuals to purposefully influence the performance of the algorithmic system. Yet although such a literacy and knowledge of algorithmic systems is key to exerting control over algorithms it also entails that individuals adopt a specific way of looking at and knowing themselves, as the following section argues.

3 Knowing the machine and knowing oneself

3.1 On reciprocity in technology use

That individuals need to accommodate algorithmic systems when trying to make them work for their purposes has been discussed in various contributions studying the efforts that individuals undertake to figure out the workings of algorithms, such as content filters of social media (e.g. Gillespie 2017; Bucher 2018; Klawitter and Hargittai 2018; Cotter 2019). Yet such attempts of gaming algorithms usually entail that individuals conform to the (presumed) logic of the algorithmic system. An instructive example is the study by Kear (2017), which shows how people may try to systematically produce behavioral data traces in order to be perceived more favorably by credit-scoring algorithms. This example illustrates how individuals may try to use algorithmic systems to their benefit by adapting to how they are perceived through the algorithmic system.

In such cases, however, users have no direct control over algorithmic systems, which means that they cannot escape the kind of 'game' predefined by others. Yet a similar dialectic even emerges if individuals are given control over how an algorithmic system performs based on the algorithmic literacy described above. There is a general element of reciprocity in terms of having to accommodate the technologies we use. As a basic requirement that marks all uses of technology, this is not per se problematic: we always need to adapt ourselves, our senses and bodily capacities and even habits to the physical and functional properties of our tools (Harris and Taylor 2005, pp. 9–10). Research from phenomenologically informed media studies has foregrounded this aspect and shown how the use of technology demands certain forms of habituation and entails specific embodied forms of knowing (Highmore 2011; Nansen et al. 2014; Parisi et al. 2017; Richardson and Hjorth 2017). Competently using a smartphone and its touchscreen, for instance, involves certain ways of directing one's senses and attention at the device, and learning and habituating certain movements and gestures to interact with it (Nansen 2020). Only once corresponding habits have been integrated into embodied knowledge can one use the device without having to think about what one is doing (O'Neal Irwin 2016, pp. 13–14).

While this habituation to our technologies is a general requirement, it can also gain moral relevance where the exigencies it entails have a constraining and disciplining effect. As Nansen et al. (2014, pp. 8–9) state, the use of interfaces and controllers even of digital media that are supposed to leverage natural gestures may require a 'reorganization of bodily movement that challenges the concept of naturalistic interaction'. In a similar vein, Finn (2017, p. 60) has pointed out that voice-controlled personal assistants are a tool that individuals can use to better manage everyday tasks and to obtain relevant information; yet they also demand of users that they learn to formulate queries in such a way that the application can 'understand' it. The technology thus trains users to interact with it in a fashion that represents a limitation of their expressive capacities.

This dialectic can similarly appear where users configure ADM systems that provide a service because the kind of literacy this demands is not just an understanding of the technology but also comprises a specific way of knowing oneself. This has to do with the fact that ADM systems rendering some service are less like conventional tools, like a hammer, but involve a delegation to an agent that, by design, incorporates certain objectives. Algorithmic systems that operate as recommender systems or personal assistants are like artificial agents serving their users and as such, they can be configured so that they calculably realize predefined objectives and trade-offs between them. However, even if the behavior of an ADM system can be customized by those to whom a service is provided, this does not avoid the dialectics described above. ADM systems are supposed to extend our control over the world, help us in leading our lives, and offer us greater convenience, but purposefully using the technology requires accommodating it: the user must view her own goals and values in a way that makes them formalizable and in this sense 'understandable' to the machine.

3.2 The operationalization challenge

Where individuals are enabled to alter aspects of an ADM system that affect its performance (i.e. central parameters) they can aim to find a configuration of the system that best corresponds to their preferences. The main challenge that individuals face in that situation is one of operationalization, of translating their goals and values into the ADM system's operations. To tackle this task, they must form an understanding of what the performance of the ADM system must look like if it is to conform to the user's own personal goals and values. They thus must explicitly ask themselves what good decision-making is to them, in terms of their own goals and values, when realized by the system.

This potentially intricate challenge can be elucidated by looking again at the example of online content filters. If designed accordingly, an algorithmic system could allow for personalizing its filtering process. Users could be enabled through an interface or dashboard to weight different forms of content, such as topics and other content characteristics, like diversity and negativity versus positivity. Clearly, this requires that the functionality of the algorithmic system makes possible this kind of personalization by registering how it performs regarding those various aspects. If this functionality is provided, however, users could substantially customize their experience in line with their needs and demands (Harambam et al. 2019).

In other cases, users might configure ADM systems by specifying what kind of prediction errors carry greater weight for them. This issue has been illustrated with regard to scheduling assistants which detect meeting requests from e-mails (Kocielnik et al. 2019). Even with the same level of overall accuracy, ADM system configurations can vary with regard to different error rates: It can create *false positives* (mails wrongly classified as positive), and *false negatives* (mails falsely classified as negative, i.e. overlooked meetings), and the ratio of these errors can be altered to some degree. Users then must decide whether they would rather have fewer false positives at the cost of more false negatives or vice versa. The question is thus: which attainable distribution of classification outcomes and errors produced by an ADM system is preferable to a person? Importantly, an ADM system will always produce some distribution of outcomes that can be linked to quantifiable parameters and performance metrics. Configuring the algorithmic system differently means the system will perform differently, including with respect to the decision errors it produces. Once a person is given control over this algorithm behavior and performance, she cannot *not* choose a specific algorithm performance. If she has the possibility to alter the performance of a recommender system, she must settle with a certain configuration because even if one simply accepts some default setting, one has opted for certain quantifiable trade-offs between decision errors and goals optimized by the system.

And various such trade-offs may exist. In the example of social media content selection, the task of translating one's values into the operations of the algorithmic system personalizing the presented content could also extend more generally to the way in which the system filters content and makes recommendations. What would the ideal diet of online content look like, i.e. what is a suitable level of polarization, how much diversity should there be in the news, how important is popularity etc.? As a person is enabled to exert control over such performance criteria of the ADM system, the more she would need to gain an understanding of what her values and preferences are exactly and how they can be expressed in the criteria and performance metrics of an ADM system.

Users will hardly have an a priori understanding about which system design best accommodates their objectives. They would have to play with different settings and find a configuration they are comfortable with—which essentially means testing themselves: who am I expressed through the parameters that guide the operations of the algorithmic system? By probing herself in this fashion a user can determine which configuration of the ADM system and which weighting of different criteria and outcomes reflects her preference and she is comfortable with.

Hence, as the example of online news content provision illustrates, dealing with the operationalization problem when exerting control over ADM systems compels users to express in a formalizable way what good decision-making means to them. They must think about and possibly reassess their own dispositions in a way that accommodates the operating mode of the algorithmic system. Hence, although the idea is to achieve a purposeful use of the technology according to human ends, this practice risks contributing further to what Winner (1980, p. 123) described as "the adaptation of human ends to technical means". Put differently, individuals may gain control over algorithms, but at the price of adapting to the way in which they are perceived through the algorithmic system's quantifying lens. It is in this sense that exerting control over algorithms can become a double-edged sword, as the following section will argue more in detail.

3.3 Ambivalence of user control over algorithm performance

As users try to control the performance of an ADM system, e.g., tune parameters of a recommender system or configure priorities in personalized social media content filters, they also make themselves knowable—to the ADM system *and* to themselves—in a specific way: Trying to express one's own dispositions in relevant metrics and parameters of an ADM system's configuration effectively amounts to a sort of psychometric test that users administer to themselves. They learn to see themselves through such metrics and to quantify their dispositions, thus being prompted to adopt what can be called a "computational mode of knowing, being, and doing" (Gilmore 2016, p. 2535). Users thereby engage in a form of self-disclosure that is mediated by, i.e. takes place through, the interaction with an algorithmic system that intervenes into the formation of their personality and identity. Crucially, this occurs under a certain formatting that is imposed by the operating logic of algorithmic systems—which performs an optimization task guided by criteria that can be variably weighted.

In this sense, configuring algorithmic systems to avoid undesirable biases and to make them correspond to one's preferences mirrors practices that have been discussed in a literature on surveillance through monitoring and tracking devices, as used, e.g., in health and fitness. These technologies have been described as devices for disciplining the self and reconfiguring the concept of selfhood through quantifying the self and the body (Lupton 2016). Specifically, these tracking technologies can produce normalizing effects as they incorporate and establish certain norms and ideals (such as body ideals) to which people compare themselves in a quest for self-optimization (Sanders 2017). This means that individuals become measured, quantified and thus made commensurable as they become exposed to a specific form of subjectivation in which they follow the imperative of "know thyself", but according to ideas, objectives, and a formatting chosen by others (Couldry and Mejias 2019, p. 170).

It is important to note that users are not simply the passive victims of these forms of subjectification and guidance. Rather, it has been noted that users can appropriate self-tracking tools in various, including creative ways that allow them to enact their own values (Sharon 2017; Bergen and Verbeek 2020). People can also purposefully use the technology to achieve a critical, reflective relation to themselves and find new forms of knowing that are potentially beneficial and empowering—at least if this is the explicit goal of engaging in such practices (Gilmore 2016, p. 2534; Bergen and Verbeek 2020, p. 10). These arguments can also be applied to algorithm-based services and the attempt to translate one's own dispositions into their performance. In this process, users can clearly discover something about themselves and, e.g., arrive at more considered preferences. One should note, however, that individuals could equally make use of increased control over algorithms to merely reinforce existing views. Especially in the context of news consumption, a person might tailor an algorithm to provide more one-sided information than would otherwise be provided—with the possible result of becoming more entrenched in extremist views. The person in question may thus better get what she wants, but this may not be in her own—and society's—best interest. It may ultimately be interesting for only very few users to use and configure algorithmic systems specifically to reflect upon one's own values.

Also, knowing oneself is not the primary goal when trying to mitigate possible undesired biases of an algorithm. Rather, making oneself the object of a specific kind of knowledge is a price to be paid. Already on the verge to the digital era and with a view to electronic mass media, Baudrillard (2017 [1976]) described a related kind of reciprocity. Taking up McLuhan's idea of the medium being the massage, Baudrillard (2017, pp. 82–86) argued that the use electronic media had a tactile quality as they entailed an incessant probing of individuals. While these draw on the media to read and decode the world around them, they are 'by the same token [themselves] constantly selected and tested by the medium' (Baudrillard 2017, p. 86). With the advent of ADM systems in information societies, these thoughts have become perhaps more relevant than ever. What can be seen as a form of user control over algorithm-based services may equally serve to probe and "datafy" individuals under a specific gaze. This is important because user control would still take place within larger information asymmetries and power structures.

These are sustained through business models that depend on creating value based on a data extraction which more and more extends into people's lives in order to quantify and commodify social relations and individuals' dispositions (Couldry and Mejias 2019). Hence, under the current conditions of "surveillance capitalism" (Zuboff 2019), giving users more control by letting them express their preferences in parameters of algorithms is likely to simply extend already existing practices of probing individuals. Not only are individuals tracked and profiled on an unprecedented scale for the purpose of processing their data and using the actionable insights gained from this activity, but even when exercising control over ADM systems by configuring them to their linking, users would still also be tested by these systems: as they probe their own dispositions to discover what configuration of an ADM system and what formalizable value trade-offs they feel comfortable with. Extending user control over recommender systems may therefore well mean that they operate as psychometric devices that increase existing trends of quantifying data subjects.

This reciprocal relationship can be fleshed out further by taking up again the phenomenological accounts on the role of habituation in technology and media use cited above. As Highmore has noted (2011, pp. 123–126), competently interacting with technology demands a habituation through which the material and cultural properties embedded in a technological artefact become internalized by their users. Indeed, as has been argued with a view to digital media, even if one is seemingly in control over a technological artefact, interaction with it is still governed by 'a grammar of interaction that follows well-defined modes of expression and navigation' (Nansen et al. 2014, p. 8). Regarding ADM systems, this reciprocal relationship takes a specific form because aligning an ADM system with one's goals and values to mitigate undesired biases involves a process of operationalization that is also directed at the dispositions of the self. The kind of algorithmic literacy involved in this process is therefore as much a literacy of the self as it is a literacy of algorithmic systems.

To illustrate how users direct computational thinking at themselves when trying to control algorithm behavior, one can look to insights obtained from stakeholder involvement in algorithm design. These cases differ from the example of content filter algorithms treated further above as they do not concern algorithmic systems which directly realize a service for consumers or citizens as primary users. After all, citizens will never directly operate, e.g., an algorithm for assessing offenders' recidivism of risk themselves. The challenge of exerting control in the examples below is, however, similar: citizens are those in whose name and in whose public interest the government may employ such algorithmic systems and they would decide which goals an algorithmic system realizes to what degree for society and what trade-offs should be made. In this sense, they are comparable to settings in which consumers directly interact with a recommender system.

Furthermore, participatory algorithm design, too, demands an algorithmic literacy that enables participants to deal with the operationalization challenge of translating their goals and values into the system. As Zhu et al. (2018, p. 20) note in the context of the challenges of co-designing algorithmic systems, it is crucial 'to educate an algorithmically literate society and promote "algorithmic thinking"'. Again, however, this thinking will also be directed at the self because stakeholders are called upon to find out what design and performance of the ADM system they feel comfortable with and reflects their preferences when expressed in parameters or metrics of the ADM system. The empirical study by Lee et al. (2019) with its elaborate framework for the participatory design of ADM systems is illustrative in this regard. Their aim is to formulate a 'method for directly involving end-users or stakeholders of algorithmic services in determining how the algorithms should make decisions'

(Lee et al. 2019, p. 6) and they show the viability of their framework with a stakeholder-designed application used to distribute expiring food from donors among non-profit organizations.

The algorithmic system in this context works to manage the workload of transporting the food to the different organizations while achieving a fair distribution (Lee et al. 2019, p. 9). As one method to determine what participants saw as fair, the authors asked respondents to explicitly state the weights of relevant criteria and thereby to formulate explicit decision rules (e.g. rating the poverty rate of a district twice as important as travel time). Notably, in the process of preference elicitation, respondents grappled with the task of translating their values into the ADM system. They also partly stated that performing evaluation of pairwise comparisons helped reevaluate and consolidate their beliefs (Lee et al. 2019, pp. 9–10). Indeed, the evaluative task that respondents performed also was a form of testing: more than merely taking a survey, people underwent a process of assessing their own values. They were prompted to probe their own decision standards (Lee et al. 2019, p. 22), including by uncovering seeming inconsistencies in their standards and evaluations: '[S]ome participants commented that they felt like they were applying internal rules inconsistently […] Explicitly specifying scores for each feature helped them reconcile their conflicting beliefs' (Lee et al. 2019, p. 9). Hence, it is notable that not only were respondents' values registered, but they were also probed regarding how their dispositions can be made explicit, consistent, and quantified via decision-rules.

In a similar vein, the study by Yu et al. proposes an interface that allows designers and stakeholders to visualize and explore trade-offs regarding algorithmic performance and to help them "select specific models that are consistent with their needs and values" (Yu et al. 2020, p. 2). Taking the example of recidivism prediction, the authors let participants use a custom-made interface that showed them how different algorithm configurations led to different ways in which false positives (false predicted as recidivism) are traded off against false negatives (falsely predicted as no recidivism). The interface is supposed to help people deal with this problem of trade-offs—which generally exists where several objectives are to be optimized –and to find an algorithm configuration that best corresponds to their own goals and values. In this process, individuals effectively learned to give numeric weights to false positives and false negatives in the prediction of recidivism—notably, an area in which the commensurability of prediction errors is at least debatable.

These examples illustrate an ambivalence in attempts to align ADM system performance with one's own preferences mentioned further above: On the one hand, the described process of preference elicitation may help participants to see themselves differently and arrive at more reflective and considered preferences. One the other hand, relating the performance and design of an algorithm to their own dispositions demands that they also make themselves an object of knowledge and learn to formalize their own decision-making preferences such that they can be incorporated in an algorithmic system.

4 Conclusion

The more people come to rely on algorithm-based services in their everyday lives, the more acute does the issue of possible undesired biases become. A direct solution for dealing with this issue would seem to lie in giving recipients of algorithm-based services control over their performance such that they can directly redress such biases. Design features which allow users to configure and tune how an algorithmic system performs—as discussed for recommender systems and online content filters—could serve that purpose and help mitigate undesired biases. However, this way of purposefully using the technology by aligning it with one's goals and values is not the straightforward solution that it may appear. Aligning the performance of the system with one's own preferences demands a certain kind of literacy that is more than merely a technical understanding of algorithmic systems. Rather, it also entails making oneself knowable in a specific fashion through a testing and interrogating of one's own preferences to determine which configuration of the algorithmic system best suits users: users make themselves knowable, including to themselves, in a way that is compatible with the algorithmic processing through adopting a perspective under which one's dispositions and aspects of one's personality appear as formalizable. It is in this sense that control over algorithms is ambivalent and can be ethically problematic: what is intended as a form of giving more control to users can easily extend and amplify an already existing tendency toward a "datafied" self that is measured, tracked, and probed (Couldry and Mejias 2019). Hence, on the path toward an 'algorithmic society' (Pasquale 2017) more user control over algorithm behavior is not per se human-centric or human compatible.

Addressing the described dialectics in exerting control over algorithmic systems may require an education for an algorithmic literacy that sensitizes individuals to the reciprocal relationship discussed above. They would have to become aware of the ambivalent nature of algorithmic literacy so that they can decide to resist it where they deem it necessary. Another solution could lie in creating AI systems—"ethics bots"—that read and learn a person's values and moral preferences from large number of behavioral observations (Etzioni and Etzioni 2017, pp. 413–414). This would occur unobtrusively, meaning that people do not have to quantify themselves according to the performance criteria of an ADM system. However, they would still become "known" to the machine, and a further, more fundamental problem remains with the idea of an "ethics bot": If a person regularly gives in to desires that she would rather not indulge, an AI system may learn to better evoke and satisfy these first-order wants rather than helping that person to better live according to her second-order wants. Hence, the challenge would be to design algorithmic systems that can learn one's preferences, including second-order preferences regarding what users should want (on this, see Russell 2020) and thus learn to become akin to human fiduciaries that adhere to relevant goals and values of those they serve (Mittelstadt 2019). Clearly, this would require much more sophisticated artificial agents than exist to date. In any case, as algorithmic systems increasingly augment and partly automate processes that would otherwise require cognitive abilities—like machines in previous industrial revolutions automated physical processes—it becomes important to think about how they can be designed in ways that accommodate humans rather than the other way around—and thus do not require that people subject themselves to a narrow and potentially limiting mentality of algorithmic thinking as the price for aligning an algorithmic system's performance with the goals and preferences of their users.

Acknowledgements I would like to thank the anonymous reviewers for their thoughtful comments and suggestions.

Funding Open Access funding enabled and organized by Projekt DEAL. The author discloses receipt of the following financial support for the research, authorship, and/or publication of this article: this research has been conducted within the project "Deciding about, by, and together with algorithmic decision-making systems", funded by the Volkswagen Foundation.

Declarations

Conflict of interest The author declares that there exists no conflict of interest.

References

Ananny M, Crawford K (2018) Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc 20:973–989. https://doi.org/10.1177/1461444816676645

Baker JJ (2017) Beyond the Information Age: The Duty of Technology Competence in the Algorithmic Society. S C Law Rev 557–578

Bakke A (2020) Everyday Googling: Results of an Observational Study and Applications for Teaching Algorithmic Literacy. Comput Compos 102577. https://doi.org/10.1016/j.compcom.2020.102577

Barocas S, Selbst AD (2016) Big data's disparate impact. Calif Law Rev 104:671–732

Baudrillard J (2017) Symbolic exchange and death, Revised edition. Sage Ltd, Thousand Oaks, CA

Bergen JP, Verbeek P-P (2020) To-do is to be: foucault, levinas, and technologically mediated subjectivation. Philos Technol. https://doi.org/10.1007/s13347-019-00390-7

Binns R (2017) Algorithmic accountability and public reason. Philos Technol. https://doi.org/10.1007/s13347-017-0263-5

Bryson JJ, Theodorou A (2019) How Society Can Maintain Human-Centric Artificial Intelligence. In: Toivonen M, Saari E (eds) Human-Centered Digitalization and Services. Springer Singapore, Singapore, pp 305–323

Bucher T (2018) If...then: algorithmic power and politics. Oxford University Press, New York

Burrell J, Kahn Z, Jonas A, Griffin D (2019) When users control the algorithms: values expressed in practices on twitter. Proc ACM Hum-Comput Interact 138:1–20

Calzada-Prado J, Marzal MÁ (2013) Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. Libri 63:. https://doi.org/10.1515/libri-2013-0010

Cotter K (2019) Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. New Media Soc 21:895–913. https://doi.org/10.1177/1461444818815684

Cotter K, Reisdorf B (2020) Algorithmic knowledge gaps: a new dimension of (digital) inequality. Int J Commun 14:745–765

Couldry N, Mejias UA (2019) The costs of connection: how data is colonizing human life and appropriating it for capitalism. Stanford University Press, Stanford, California

D'Ignazio C, Bhargava R (2015) Approaches to building big data literacy. Bloom Data Good Exch Conf Sept 2015 N Y Citiy 1–6

Diakopoulos N (2014) Algorithmic accountability reporting: on the investigation of black boxes. Tow Cent Digit Journal Publ. https://doi.org/10.7916/d8zk5tw2

Edwards L, Veale M (2017) Slave to the algorithm: why a right to an explanation is probably not the remedy you are looking for. Duke Tech Rev 16:18–84

Eslami M, Karahalios K, Sandvig C, et al (2016) First I "Like" It, then I Hide It: Folk Theories of Social Feeds. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp 2371–2382

Etzioni A, Etzioni O (2017) Incorporating ethics into artificial intelligence. J Ethics 21:403–418. https://doi.org/10.1007/s10892-017-9252-2

Felzmann H, Fosch-Villaronga E, Lutz C, Tamò-Larrieux A (2020) Towards transparency by design for artificial intelligence. Sci Eng Ethics 26:3333–3361. https://doi.org/10.1007/s11948-020-00276-4

Finn E (2017) What algorithms want: imagination in the age of computing. MIT Press, Cambridge

Gillespie T (2017) Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem. Inf Commun Soc 20:63–80. https://doi.org/10.1080/1369118X.2016.1199721

Gilmore JN (2016) Everywear: the quantified self and wearable fitness technologies. New Media Soc 18:2524–2539. https://doi.org/10.1177/1461444815588768

Gran A-B, Booth P, Bucher T (2021) To be or not to be algorithm aware: a question of a new digital divide? Inf Commun Soc 24:1779–1796. https://doi.org/10.1080/1369118X.2020.1736124

Harambam J, Bountouridis D, Makhortykh M, van Hoboken J (2019) Designing for the better by taking users into account: a qualitative evaluation of user control mechanisms in (news) recommender systems. In: Proceedings of the 13th ACM Conference on Recommender Systems. ACM, Copenhagen Denmark, pp 69–77

Hargittai E, Gruber J, Djukaric T et al (2020) Black box measures? How to study people's algorithm skills. Inf Commun Soc 23:764–775. https://doi.org/10.1080/1369118X.2020.1713846

Harper FM, Xu F, Kaur H, et al (2015) Putting Users in Control of their Recommendations. In: Proceedings of the 9th ACM Conference on Recommender Systems. ACM, Vienna Austria, pp 3–10

Harris JL, Taylor PA (2005) Digital matters: theory and culture of the matrix. Routledge, London, New York

Helberger N, Pierson J, Poell T (2018) Governing online platforms: From contested to cooperative responsibility. Inf Soc 34:1–14. https://doi.org/10.1080/01972243.2017.1391913

Highmore B (2011) Ordinary lives: studies in the everyday. Routledge, London

Hildebrandt M (2016) Law as information in the era of data-driven agency: law as information. Mod Law Rev 79:1–30. https://doi.org/10.1111/1468-2230.12165

Holzinger A, Kieseberg P, Weippl E, Tjoa AM (2018) Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI. In: Holzinger A, Kieseberg P, Tjoa AM, Weippl E (eds) Machine learning and knowledge extraction. Springer International Publishing, Cham, pp 1–8

Hsu S, Vaccaro K, Yue Y, et al (2020) Awareness, Navigation, and Use of Feed Control Settings Online. CHI 20 Proc 2020 CHI Conf Hum Factors Comput Syst 1–12

Kear M (2017) Playing the credit score game: algorithms, 'positive' data and the personification of financial objects. Econ Soc 46:346–368. https://doi.org/10.1080/03085147.2017.1412642

Klawitter E, Hargittai E (2018) "It's like learning a whole other language": the role of algorithmic skills in the curation of creative goods. Int J Commun 12:3490–3510

Kocielnik R, Amershi S, Bennett PN (2019) Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, Glasgow Scotland Uk, pp 1–14

Krafft TD, Zweig KA, König PD (2020) How to regulate algorithmic decision-making: a framework of regulatory requirements for different applications. Regul Gov (online First). https://doi.org/10.1111/rego.12369

Kroll JA, Huey J, Barocas S et al (2017) Accountable algorithms. univ pa. Law Rev 165:633–705

Lanzing M (2018) "Strongly Recommended" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies. Philos Technol. https://doi.org/10.1007/s13347-018-0316-4

Lee MK, Psomas A, Procaccia AD et al (2019) WeBuildAI: participatory framework for algorithmic governance. Proc ACM Hum-Comput Interact 3:1–35. https://doi.org/10.1145/3359283

Lepri B, Oliver N, Letouzé E et al (2018) Fair, transparent, and accountable algorithmic decision-making processes: the premise, the proposed solutions, and the open challenges. Philos Technol 31:611–627. https://doi.org/10.1007/s13347-017-0279-x

Lloyd A (2019) Chasing Frankenstein's monster: information literacy in the black box society. J Doc 75:1475–1485. https://doi.org/10.1108/JD-02-2019-0035

Lupton D (2016) The diverse domains of quantified selves: self-tracking modes and dataveillance. Econ Soc 45:101–122. https://doi.org/10.1080/03085147.2016.1143726

Malgieri G (2019) Automated decision-making in the EU member states: the right to explanation and other "suitable safeguards" in the national legislations. Comput Law Secur Rev 35:1–26. https://doi.org/10.1016/j.clsr.2019.05.002

Malgieri G, Comandé G (2017) Why a right to legibility of automated decision-making exists in the general data protection regulation. Int Data Priv Law 7:243–265. https://doi.org/10.1093/idpl/ipx019

Matheus R, Janssen M, Maheshwari D (2020) Data science empowering the public: data-driven dashboards for transparent and accountable decision-making in smart cities. Gov Inf Q 37:1–9. https://doi.org/10.1016/j.giq.2018.01.006

Metz R (2021) Facebook's success was built on algorithms. Can they also fix it? CNN

Mittelstadt B (2019) Principles alone cannot guarantee ethical AI. Nat Mach Intell 1:501–507. https://doi.org/10.1038/s42256-019-0114-4

Mittelstadt BD, Allo P, Taddeo M et al (2016) The ethics of algorithms: mapping the debate. Big Data Soc 3:1–21. https://doi.org/10.1177/2053951716679679

Nansen B (2020) A touchscreen media habitus. Young Children and Mobile Media. Springer International Publishing, Cham, pp 53–69

Nansen B, Vetere F, Robertson T et al (2014) Reciprocal habituation: a study of older people and the kinect. ACM Trans Comput-Hum Interact 21:1–20. https://doi.org/10.1145/2617573

O'Neal Irwin S (2016) Digital media: human-technology connections. Lexington Books, Lanham

Parisi D, Paterson M, Archer JE (2017) Haptic Media. Stud New Media Soc 19:1513–1522. https://doi.org/10.1177/1461444817717518

Pasquale F (2017) Toward a fourth law of robotics: preserving attribution, responsibility, and explainability in an algorithmic society. Ohio State Law J 78:1243–1255

Pasquale F (2015) The Black box society: the secret algorithms that control money and information. Harvard University Press, Cambridge, Massachusetts London, England

Rainie L, Anderson J (2017) Code-dependent: pros and cons of the algorithm age. Pew Research Center, Washington, D.C.

Reisman D, Schultz J, Crawford K, Whittaker M (2018) Algorithmic impact assessments: a practical framework for public agency accountability. New York University, New York, AINow Institute

Richardson I, Hjorth L (2017) Mobile media, domestic play and haptic ethnography. New Media Soc 19:1653–1667. https://doi.org/10.1177/1461444817717516

Robinson HM (2009) Emergent computer literacy: a developmental perspective. Routledge, New York

Russell SJ (2020) Human compatible: AI and the problem of control. Penguin Books, London

Samek W, Müller K-R (2019) Towards explainable artificial intelligence. In: Samek W, Montavon G, Vedaldi A et al (eds) Explainable AI: interpreting, explaining and visualizing deep learning. Springer International Publishing, Cham, pp 5–22

Sander I (2020) What is critical big data literacy and how can it be implemented? Internet Policy Rev 9:. https://doi.org/10.14763/2020.2.1479

Sanders R (2017) Self-tracking in the digital era: biopower, patriarchy, and the new biometric body projects. Body Soc 23:36–63. https://doi.org/10.1177/1357034X16660366

Saurwein F, Just N, Latzer M (2015) Governance of algorithms: options and limitations. info 17:35–49. https://doi.org/10.1108/info-05-2015-0025

Sharon T (2017) Self-tracking for health and the quantified self: re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare. Philos Technol 30:93–121. https://doi.org/10.1007/s13347-016-0215-5

Vaccaro K, Huan D, Eslami M, et al (2018) The Illusion of Control: Placebo Effects of Control Settings. CHI 18 Proc 2018 CHI Conf Hum Factors Comput Syst 1–13

van Drunen MZ, Helberger N, Bastian M (2019) Know your algorithm: what media organizations need to explain to their users about news personalization. Int Data Priv Law 9:220–235. https://doi.org/10.1093/idpl/ipz011

Wachter S, Mittelstadt B, Floridi L (2017) Transparent, explainable, and accountable AI for robotics. Sci Robot 2:eaan6080. https://doi.org/10.1126/scirobotics.aan6080

Winner L (1980) Do artifacts have politics? Daedalus 109:121–136

Yeung K (2017a) Algorithmic regulation: a critical interrogation: algorithmic regulation. Regul Gov 12:505–523. https://doi.org/10.1111/rego.12158

Yeung K (2017b) ‘Hypernudge’: big data as a mode of regulation by design. Inf Commun Soc 20:118–136. https://doi.org/10.1080/1369118X.2016.1186713

Ytre-Arne B, Moe H (2021) Folk theories of algorithms: understanding digital irritation. Media Cult Soc 43:807–824. https://doi.org/10.1177/0163443720972314

Yu B, Yuan Y, Terveen L, et al (2020) Keeping Designers in the Loop: Communicating Inherent Algorithmic Trade-offs Across Multiple Objectives. In: Proceedings of the 2020 ACM Designing Interactive Systems Conference. ACM, Eindhoven Netherlands, pp 1245–1257

Zarouali B, Helberger N, De Vreese CH (2021) Investigating algorithmic misconceptions in a media context: Source of a new digital divide? Media Commun 9:134–144. https://doi.org/10.17645/mac.v9i4.4090

Zhu H, Yu B, Halfaker A, Terveen L (2018) Value-sensitive algorithm design: method, case study, and lessons. Proc ACM Hum-Comput Interact 2:1–23. https://doi.org/10.1145/3274463

Zuboff S (2019) The age of surveillance capitalism: the fight for the future at the new frontier of power. Profile Books, London

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.