Kölner Zeitschrift für Soziologie und Sozialpsychologie

Why Organization Matters in "Algorithmic Discrimination"

Rena Schwarting · Lena Ulbricht

Abstract Research into "algorithmic discrimination" has largely dismissed the fact that algorithms are often developed and used by organizations. In this article, we show that organizational sociology can contribute to a more nuanced perspective on "algorithmic decision-making." Drawing on the concept of decision premises, we differentiate between various formal structures, particularly between different decision programs (conditional and purposive). This allows us to challenge two key assumptions, namely that human decision-makers rely heavily on algorithmically generated recommendations and that discrimination against protected groups needs to be solved mainly at the level of code.

We identify the usefulness of distinguishing between conditional and purposive decision programs via a case study centered on the legal context: the risk assessment software "Correctional Offender Management Profiling for Alternative Sanctions" (COMPAS) that is employed in the US criminal justice system to inform judicial personnel about the recidivism risk of defendants. By analyzing the organizational structures, according to which the COMPAS score is formally and informally embedded in courts, we point out that the score represents an ambiguous and redundant

R. Schwarting (✉)
WZB Berlin Social Science Center
Reichpietschufer 50, 10785 Berlin, Germany
E-Mail: rena.schwarting@wzb.eu; rena.schwarting@unisg.ch

University of St. Gallen
St. Gallen, Switzerland

L. Ulbricht
WZB Berlin Social Science Center
Reichpietschufer 50, 10785 Berlin, Germany
E-Mail: lena.ulbricht@wzb.eu

Weizenbaum Institute for the Networked Society
Berlin, Germany

information source for judges. The practice of minimizing the relevance of the score and decoupling it from the legal reasoning backstage particularly reflects the professional decision autonomy of judges, which is inherent in the legal system. The core finding of our approach is that strategies to reduce discrimination should not only scrutinize data quality or the statistical model but also consider the specific forms, functions, and consequences of the organizational structures that condition the ways in which discriminatory differences may or may not be (re)produced.

Warum Organisationen einen Unterschied bei „algorithmischer Diskriminierung" machen

Zusammenfassung Die Forschung über „algorithmische Diskriminierung" hat bislang weitgehend außer Acht gelassen, dass Algorithmen von Organisationen entwickelt und eingesetzt werden. Durch den Rückgriff auf die Unterscheidung von formalen und informalen Organisationsstrukturen im Allgemeinen und das Entscheidungsprämissenkonzept im Besonderen, so unser Argument, können zwei zentrale Thesen über den Einsatz von Software in Organisationen relativiert werden: Organisationstheoretisch zu problematisieren sind zum einen die Annahme, dass algorithmisch generierte Daten bruchlos in Entscheidungsstrukturen übernommen würden, und zum anderen die Behauptung, dass diskriminierende Folgen hauptsächlich auf der Ebene des Software-Codes zu lösen seien.

Den Erkenntnisgewinn einer organisationssoziologischen Perspektive verdeutlichen wir exemplarisch anhand einer Fallstudie aus dem Rechtsbereich: der Risikobewertungssoftware COMPAS, die in der US-amerikanischen Strafjustiz zur Bewertung des Rückfallrisikos von Angeklagten eingesetzt wird. Durch die Analyse der Entscheidungsstrukturen, nach denen der COMPAS-Score inter- und intraorganisatorisch eingebettet ist, zeigen wir auf, dass der Score eine mehrdeutige und redundante Information für richterliche Entscheidungen darstellt. Die in empirischen Studien beobachtete Praxis, dass Richterinnen und Richter die Relevanz des Scores minimieren und ihn bei der juristischen Entscheidungsfindung auf der Hinterbühne abkoppeln, spiegelt insbesondere ihre professionelle Entscheidungsautonomie als formale Organisationsmitglieder von Gerichten wider. Die wichtigste Erkenntnis unseres Beitrags ist, dass Strategien zur Verringerung von diskriminierenden Folgen nicht nur die Qualität der Daten und der statistischen Modelle untersuchen sollten. Vielmehr gilt es, insbesondere die spezifischen, eigenlogischen Formen, Funktionen und Folgen der Organisationsstrukturen zu verstehen, welche die Art und Weise bedingen, wie diskriminierende Unterschiede (re-)produziert werden können oder auch nicht.

1 Introduction

"There's software used across the country to predict future criminals. And it's biased against blacks." This is the conclusion of an investigative news story published by ProPublica (Angwin et al. 2016). The story was based on a research report that ignited a major international controversy about the question of whether risk assessment software is more racially biased than judges. The debate prompted an academic response in the newly emerged fields of fairness, accountability, and transparency in machine learning, critical algorithm studies, and AI ethics and data justice, but much less so in areas that study organizational decision-making. The aim of this article is to identify the benefits of adding an organizational perspective to the study of what is called "algorithmic discrimination."

"Algorithmic systems" have been established for a long time in history, although in the form of regulations or recipes for structuring almost every human interaction by minimizing (human) interpretation (Daston and Galison 2007). In the last few decades, "algorithmic systems" have often been simply automated standardized risk instruments, in the form of ratings and rankings, for example, used in financial institutions (Carruthers and Ariovich 2010; Kette 2008; Schwarting 2010; Sinclair 2008; Strulik and Willke 2007), to "rationally" calculate and categorize uncertain behavior in financial markets (Poon 2007; Rona-Tas and Hiss 2010), or in public organizations, such as universities, hospitals, and prisons (for the UK, see Mennicken 2013).

Consistent with the common definition in computer science, we see an algorithm as "an abstract, formalized description of a computational procedure" (Dourish 2016). Although the term "algorithm" is a central point of reference in current studies, we argue that it is both over- and undergeneralized for the purposes of sociological research. A single algorithm or its linkage in a computer program does not represent a social problem per se as there is no a priori dichotomy; there is not even a hierarchy between algorithms and social interaction in general or between algorithms and organizational decision-making in particular. Rather, and in line with organizational sociology, we presume that members of organizations and algorithmically generated data cooperatively interact, sometimes one before the other and sometimes one after the other; they do not do so in a clear order but according to need.

To take account of a social-constructivist perspective on algorithms that reflects their technical and social embeddedness as rules, we acknowledge that algorithms are not necessarily relevant for the structuring of social actions (Dourish 2016; Foot et al. 2014; Yeung 2018).[1] What is called "algorithmic decision-making" is therefore simply decision-making that is somehow associated with algorithms. The relationship may be loose, in the form of "recommendations," or rigid, described as "automated decision making" (ADM). In addition, we define discrimination as

---

[1] It is noteworthy that, outside computer science, the attribute "algorithmic" is often used to refer to models of automation (Dourish 2016). In these cases, the analysis does not focus on the lines of code of an algorithm or on a set of algorithms and data (often coined "algorithmic systems") but rather the social practices that are shaped by some degree of automation.

something that takes place when specific group-related differences (e.g., age, gender, race) structure communication and thereby create disadvantages for groups that are legally protected. By "algorithmic discrimination," we understand processes in which algorithmic phenomena, mostly software,[2] contribute to discriminatory effects. In this sense, discrimination related to algorithms can take many forms, as we elaborate later on.

Research on the relation between algorithms (including the "big data" generated in the process) and discrimination is highly inter-disciplinary, fragmented, and methodologically diverse, but has not yet delved deeply into sociological theory on the conditions and consequences with regard to discrimination and organizational decision-making. Relevant studies are often aimed at identifying the scope of discrimination and the criteria upon which discrimination takes place and at exploring different mechanisms that lead to "algorithmic discrimination," be they the data sets (which can be "biased"), the calculative models (which may not adequately account for heterogeneity), and the execution or interpretation of algorithmic recommendations (which can be too decontextualized, too inflexible, too selective, etc.) (Mitchell et al. 2021).

Empirical studies have identified forms of "algorithmic discrimination" against legally protected groups in search engines (Araújo et al. 2016; Caliskan et al. 2017; Lambrecht and Tucker 2016; Noble 2018; Sweeney 2013; Kay et al. 2015); on platforms such as Uber (Rosenblat et al. 2017) or Airbnb (Edelman et al. 2017; Gilheany et al. 2015); and on self-described "social networks" like Facebook (Angwin et al. 2017; Hofstra and Schipper 2018). In addition to the "algorithmic discrimination" associated with commercial platforms that privately organize, curate, and commodify large parts of both private and public exchange because of their market concentration (Dolata 2019), risk assessment software has also been characterized as biased. Prominent cases have been revealed in the credit industry (Avery et al. 2012), in criminal justice (Angwin et al. 2016; Benjamin 2019; Chouldechova 2017; Hamilton 2019a, b), in social policy (Cech et al. 2019), or in recruitment (Lowry and MacPherson 1988; Rosenblat et al. 2014).

However, these studies mainly focus on the design of digital technologies and the divergent impacts on protected groups, but not on the concrete day-to-day handling of data in actual decision-making. Such perspectives neglect the fact that, in modern societies, communication is differentiated by various forms of social order that can conflict, such as families, organizations, networks, or functional fields. These contexts vary in their constitutive structures and therefore in how political, administrative, and economic logics shape data usage. Indeed, scholars who theorize algorithms as a form of social ordering through quantification (Beckert and Aspers 2011; Heintz and Wobbe 2021; Mehrpouya and Samiolo 2016), governance by numbers (Heintz 2008), or algorithmic governance/regulation (see, for example Katzenbach and Ulbricht 2019; Yeung 2018) have occasionally pointed out that the

---

[2] Algorithms and computer programs (or software) are different entities: Programs may embody or implement algorithms, but "programs are both more than algorithms (in the sense that programs include non-algorithmic material) and less than algorithms (in the sense that algorithms are free of the material constraints implied by reduction to particular implementations)" (Dourish 2016).

implementation and effects of "big data" are likely to vary along with the organizational contexts in which they appear. Based on neoinstitutionalist theory, a few studies view "datafication" (Couldry and Yu 2018) and "algorithmification" (Gillespie 2014) as coupling mechanisms between societal subsystems (e.g., the economy and the legal system) and different types of organization (e.g., companies and courts) that might change their formal structures and logics (or not). Additionally, Caplan and Boyd (2018) have identified an "algorithmic logic," which they consider to be both a macro-trend and a mechanism through which different organizations in a field align themselves with regard to their formal structure; this results in a process of increasing "institutional homogeneity" and "isomorphism" (DiMaggio and Powell 1983; Meyer and Rowan 1977).[3]

Given that organizations and their history are essential for understanding the use of algorithms (see also Graeber 2016, p. 40), it is striking that none of these studies makes use of the rich body of organizational theory. From an organizational perspective, algorithms are, on the one hand, products and services of organizations; on the other hand, they are used by organizations to regulate their interactions with their members, clients, and (or including) other organizations. Consequently, in this article, we ask whether and how concepts from organizational theory can contribute to a differentiated perspective on algorithmic decision-making in organizations.

To flesh out our argument, we suggest a conceptual framework for assessing how organizations process algorithmic information in criminal justice and how it is used to make decisions about defendants. Our approach draws on the concepts of bounded rationality (Cohen et al. 1972, 1994) and decision programs (March and Simon, 1958; Simon 1947, 1991) elaborated in systems theory (Luhmann 1966, 2018) as well as on current studies on resistance to change in organizations (Ybema and Horvers 2017). We argue that the ways in which information generated by data-driven technologies is used in organizations and made relevant for decision-making depend on how it is integrated into organizational settings (see also Alaimo and Kallinikos 2020; Beverungen et al. 2019; Büchner 2018; Büchner and Dosdall 2021).

In a first step, we present our general argument on why organization matters in algorithmic discrimination. By conceptualizing discrimination from a sociological perspective as group-related inequalities, we point out that group-related exclusion is an inherent feature of modern societies. To specify the practices of reproducing, ignoring, or neutralizing group-related differences, we draw on organizational sociology and distinguish between the specific decision-making structures (personnel, rules/programs, and hierarchies and/or networks) of different types of organizations and their conflicting rationalities and purposes (e.g., economic, legal, or administrative logics). In particular, we differentiate between two different decision rules, namely conditional and purposive programs. Based on this framework, the third section of this article analyzes the risk assessment software "Correctional Offender Management Profiling for Alternative Sanctions" (COMPAS) in the context of US courts. We point out that the score represents an ambiguous and redundant source

---

[3] Neoinstitutionalism also provides valuable insights for gender-sensitive organizational research (Funder 2018), which, however, has not dealt with digital technologies in particular.

of information for judges owing to its limited accuracy and the pre-existing biases inscribed in the data. After identifying the decisional structures that guide how (algorithmically generated) information is (not) made relevant, we turn to the actual use of the COMPAS score by judges, which can be characterized as "resistance through compliance"—this entails strategies of open resistance, foot-dragging, and criteria-tinkering. The final section discusses the implications of our contribution and possible further research.

2 An Organization-Based Framework for "Algorithmic Decision-Making" and Discrimination

Organizational sociology studies organizations as a specific form of social order; it looks at their formal and informal structures, professional norms, and the related effects. Apart from systematizing various types of organization, this body of research examines the diverse tensions between organizational autonomy and the relation between organizations and their environment (Cohen et al. 1972, 1994; DiMaggio and Powell 1983; Luhmann 2018; Thompson 1967). We introduce two concepts from organizational theory that contribute to a deeper understanding of how protected groups' personal features may structure decision-making. First, drawing on theories of societal and social differentiation, we distinguish between different social contexts, namely families or networks, where interactions are mostly informal, and formal organizations, including courts, where interactions are both formal and informal. Second, we propose the concept of "decision premises" (March and Simon 1958; Simon 1947, 1991) and present two types of organizational decision programs that may attribute different functions to algorithmic recommendations.

Generally speaking, the use of these distinctions reveals that whether discriminatory differences, such as age, race, or gender differences, become relevant in communication relies on the expectations of an observer and on the structures of the social contexts into which modern societies are broken down. The fact that group-related differences no longer entirely determine a person's inclusion in society is a distinctive feature of modern societies. In the course of history, formal organizations were established as a novel type of social order between society-at-large and individual interactions in friendships, networks, or families. With the rise of organizations (Etzioni 1961; Perrow 1991; Zald 1990), individuals were only partially included in various organizations based on roles that were both temporal and subject to possible exclusions. A person could thus maintain multiple memberships and associated roles. However, many studies have pointed out that, although organizations formally reject discrimination based on law and equality standards, they have not been able to root it out entirely. This is because there are informal practices in organizations that are not subject to formal expectations and that enable discrimination. Put differently: Although at the formal level, group-related differences are restricted by factual considerations, informal interactions in organizations lack functional specifications. Indeed, in communication that is not framed formally, face-to-face interactions not only transmit information but also serve to express feelings, to present oneself, and to motivate contacts through the multi-

ple biological channels of human sensation. Under such conditions of physical co-presence—which is key in legal administration and court hearings dealing with life and death decisions—group-related characteristics are likely factors that structure communication and decisional heuristics in order to absorb uncertainty (Gigerenzer and Engel 2006),[4] owing to their "simple informational manageability" (Tacke 2008, p. 259). Consequently, when seeking to understand how discriminatory differences associated with automated, data-driven technologies in organizations do or do not serve as a reference for decision-making, it is important to specify the extent to which decision rules are formalized and thus given binding power.

Research into decision science has traditionally assumed a strong link between information and decisions—it regards decision-making as a one-sided, essentially rational operation in which individuals process information in a decontextualized manner. In contrast, constructivist decision theory has revealed the various forms and mechanisms of bounded rationality associated with decision-making (Simon 1991). For instance, Martha Feldman and James March (1981) have pointed out that much of the information that is gathered and communicated in organizations has little decision-making relevance: Regardless of the information available, information is often neglected and more information is requested. They explain this practice of information ignorance and overload with reference to the symbolic function of information. The information collected by organizations (e.g., in the form of risk assessments) is not necessarily obtained in "decision mode" but also in "surveillance mode," for instance, to justify and rationalize decisions ex post (Feldman and March 1981, p. 175; see also March and Shapira 1987). Hence, the relevance and collective meaning of information for decision-making cannot be taken for granted: As a product of communication information is a social thing, and researchers should consider the decision structures within a specific organization.

As previously stated, organizations link decisions to exclude group members to the nonfulfillment of their membership conditions. The expectations associated with membership are part of the formal order of an organization; and structures are relatively constant expectations of behavior. However, expectations do not fully determine communication but merely regulate it; structures thus make certain behavior more likely while discouraging alternatives.[5] According to March and Simon (1958), past decisions function as a reference for an indefinite number of future decisions, serving as unquestioned principles known as "decision premises" (Simon, 1947). Elaborating on these pioneering works in a systems-theoretical account on organizations, Luhmann distinguishes between three decision premises: communication channels (hierarchies and/or networks), decision rules/programs, and personnel. Communication channels define how and when members may formally get in touch with each other. Decision programs characterize the rules that represent the proper

---

[4] However, it is worth noting that the importance of physical characteristics in face-to-face interaction can also be reduced by other forms of (standardized) self-presentation—such as uniform clothing or depersonalized names.

[5] It is important to differentiate between decisional rules in the sense of executing predefined tasks and legal procedures as a specific type of social interaction aimed at deriving a legitimate verdict (Luhmann 1983; Schwarting 2020).

Table 1 Two types of decision programs

Type of decision program	Means	Purpose	Coupling means-purpose	Logic	Locus of "(un)fair" treatment and public accountability
Conditional	Defined	Not specified	Pre-defined	"If A, then B."	Data and algorithm
Purposive	Not specified	Defined	Not specified	"What is required to attain B?"	Entire information selection process

Source: own representation

conditions for the execution of tasks. Personnel implies that a person's cognitive competence and physical ability make a difference about the decisions that they take (Luhmann 2018, p. 210ff.).

By concentrating on the rules that orient whether and how algorithmically generated information becomes relevant for decision-making in organizations, scholars account for the fact that rules can also be understood as "depersonalized authority" in that they regulate the execution of a decision on information independently of the presence of a person-related rank. Decision programs may be further broken down into two types of rules: conditional programs and purposive programs (Luhmann, 1966, p. 36ff., 77, 2018, p. 210ff.; March and Simon 1958, p. 165ff., 185ff.). Conditional programs are strict if-then rules that define the conditions for decisions. In contrast, purposive programs draft the ends for which appropriate means are to be chosen. A famous example is "management by objectives."

We exemplify the difference between conditional and purposive programs in the table below. Note that for the sake of illustration it is schematic: Generally speaking, the two forms of executing decisions can be described in situations where A represents the means and B the purpose (Tab. 1). In the case of court decisions, A includes all information that is taken into consideration for a judge's decision, and B is a "rightful verdict" (i.e., a verdict that is legitimate, etc.). Conditional programs essentially follow the principle of "if A, then B." The means are specified and therefore called "conditions," and lead to a purpose that can change as a direct consequence of the conditions. Means and purposes are therefore rigid, or more or less tightly coupled, and human decision-makers are not held responsible for decisions.

Instead, purposive programs ask, "What means are required to attain B?". The purpose typically does not unambiguously define the action but only functions as a reference point to compare and select the appropriate means. Although the purpose is more or less pre-defined, the means are unspecified and can be chosen rather by the decision-makers, who are therefore—and in contrast to the situation in conditional programs—responsible for their decision. Here, means are unspecified; the final decision is taken based on many informational factors and not just one risk score—for example, a psychological report, the hearing, the judge's intuition, etc. Although this is often neglected in classical decision theory, the selection of information can be risky and contradictory, because purposes are abstract and thus give little indication of the exclusion of other means. Purposive programs require more ambiguity tolerance, whereas conditional programs establish an elevated level of

authority when processing information and executing decisions. Against this backdrop, where statistically "fair" software computing and modelling are important to prevent the (re-)production of discriminatory data, the structures (including personnel, rules/programs, and communication channels) according to which information is selected and considered "legitimate" for decision-making in organizations are equally crucial.

The distinction between conditional and purposive programs is central as it can be applied to the type of organization, to the specification of the organizational structures, and to "technical systems" or "classes of algorithms". At an organizational level, the scope of and relationship between conditional and purposive programs varies depending on the type of organization. Although enterprises, because of their higher decision autonomy, may alter their purposes, public administrations and courts traditionally operate via conditional programs (Luhmann 1966).

Likewise, digital technologies can be understood as decision premises defining rules for the execution of repeatable tasks. Algorithms quite perfectly materialize decision programs for the execution of decisions. The exact design, however, has already been decided within an organization. For example, classic regression models can be understood as largely conditionally programmed. In contrast, what is called "semi-supervised machine learning" can be regarded as a purposive program (composed of various conditional programs)—that is, it can be viewed as a kind of algorithm that has been pre-structured with a certain selection of means to achieve a pre-selected output. With this in mind, it is apparent that algorithms themselves do not decide on their features in contingent ways. Rather, algorithms execute previously defined rules with reference to decontextualized input data, sophisticated output functions, and formal logics.

In a similar way, decision programs help us to distinguish between two modes of human–computer interaction. The first mode concerns systems where computer-generated decisions are automatically implemented (by technical means, or by human decision-makers who systematically stick to certain recommendation information); this is often called "ADM" and would qualify as a conditional program. The second mode implies that "algorithmic recommendations" need to be interpreted by human decision-makers who take the final decision—the currently debated "human-in/on-the-loop" model that is common in many "assistance systems"; this represents a purposive program.

3 Mediating Algorithms in Organizations: The COMPAS Score in US Courts

Taking the example of the COMPAS score, we will demonstrate how the concepts of decision premises developed in the previous section can help us to understand the use of big data in organizations and its potential for (re)producing discriminatory effects.

The software COMPAS was developed in 1998 by the company Northpointe and designed to assess the likelihood of a defendant committing a crime within the next two years. According to the promotional materials, the score was introduced as

a "case management and decision support tool" to more efficiently deploy resources in largely privatized, "overloaded and crowded criminal justice systems" (Northpointe 2015, p. 2). It has been used by a number of jurisdictions, including Florida, New York, Wisconsin, and California. In addition to COMPAS, more than 60 other risk instruments are now in use throughout the United States in almost all areas of criminal justice, from preliminary proceedings to probation and conviction. Other countries, such as Canada, Australia, and several European states, are also developing similar risk assessments (for the origin and development of risk assessment systems in criminal justice, see Kehl et al. 2017; Monahan and Skeem 2016).

We have chosen COMPAS for three reasons. The first reason is that courts are organizations with a long tradition of setting up rules for documenting and processing data. The fact that courts are different from other types of organization, as they cannot modify their purpose, is often neglected. Legal public administration is—in contrast to companies—mainly structured by conditional premises that are aimed at neutralizing the effect of personal expectations on decisions. A company sees the computer as an instrument of management and of improving the quality of decision-making. This is because, in private-sector organizations, the purposes are specific, and the choice of means is variable. In the production of administrative acts; however, the result is politically and legally pre-determined. In public legal administration, various considerations must be taken into account, and these are not easily quantifiable. It may not be better or worse to use technology than it would be to perform the same task manually (Luhmann 1966, p. 16f.).

Secondly, courts are a special type of professional organization, and like (primary) schools or hospitals, they cannot choose or reject their clients but have to "wait" to take actions (Schwarting 2020). Thus, they find themselves in a difficult double role: on the one hand, they have to select and therefore produce or reproduce social inequality; on the other hand, they are supposed to be the place where previous social inequality—or society's conception of equality—can still be "corrected". Also, and in contrast to nurses, teachers, or social workers (Ackroyd 2016; Etzioni 1969), judges and lawyers are recognized as full *professions* in sociology. Unlike expert fields that encompass fuzzier occupations such as journalism, professions exercise a monopoly on their jurisdiction, including the strict control of admission to and the organization of knowledge and work by disbarring outsiders in cases of malpractice (Abbott 1988). With these privileges and their expertise, they have significant autonomy in how they categorize and diagnose situations, with serious consequences for the individual lives and death of their "clients" or "customers." For these reasons, many professions used to be protected from quantitative evaluation. However, as the COMPAS score also reveals, even professional organizations are now asked to comply with a growing number of metrics and standards (Brunsson and Jacobsson 2000).

Thirdly, and more empirically speaking, there are further aspects that make courts a rich setting for studying the regulation of big data in organizations. For one, there is the abundant evidence for the COMPAS system, its data, its calculative procedures, and its use by legal professionals. Moreover, its daily use by professionals who are supposed to have limited training in statistics, algorithmic design, and data analytics make it an intriguing example through which to examine the interplay between data

analytics and the different organizational roles and rules of courts—that, finally, represent a context that has been studied comparatively little so far in organizational research (Schwarting 2020).

3.1 Biased Data: Organizational Inscription of Social Inequalities

If information is to play a role in organizational decision-making, it needs to be considered a legitimate source. For the introduction of the COMPAS software, the company Northpointe referred to studies that assumed that statistical methods would be more accurate and objective than human judgment (Casey et al. 2014, p. 4; Northpointe 2015, p. 2). Scores would therefore counteract the possibly racially biased "subjectivities" of judges with the "objective" information of a computer-generated calculation (Angwin et al. 2016). However, before we identify the organizational structures according to which the score does (not) gain collective relevance for judicial subsumption, we will show in this section how the statistical "accuracy" and "fairness" of the score have already been challenged.

So far, COMPAS scores have been collected from over one million people. The scoring relies on a questionnaire with approximately 137 differently weighted variables that are combined in a variety of subscales. In addition to the type and number of criminal records, socio-demographic features are modelled (e.g., age, gender, drug use, education, income and employment status, residential stability, family structure, community ties, and personal attitudes).[6] A rating of 1 to 4 indicates a low risk, 5 to 7 a moderate risk, and 8 to 10 a high recidivism risk. The risk level itself is noted in a report that is sent to the judges. The results are usually shared with the defendant's attorney. However, the details of the calculative procedures are not publicly available, which Northpointe justifies with reference to trade secrecy and the risk that defendants might want to "game the system." A study that was aimed at reconstructing the COMPAS model found that the public statements made by Northpointe about the formal logic of its model—for instance, concerning the linearity of the age factor—do not hold true (Rudin et al. 2019, p. 2). Another analysis concluded that the COMPAS model does not deliver better predictions than those made by people with little or no criminal justice expertise and a model with only two determinants, age and criminal history. The latter model yielded a prediction with which the authors were satisfied (67%) and which, interestingly, was better than the one from COMPAS (65%) (Dressel and Farid 2018). These studies cast serious doubt about the accuracy of the COMPAS score and openly question the value of a large variety and quantity of input data and the underlying calculative model.

In addition, and also in contrast to its intended aim, the score has been criticized for reproducing racial bias (Angwin et al. 2016). Specifically, the analysis show that "blacks" were almost twice as likely to be classified with higher risk values as "whites," although they did not reoffend. Conversely, whites were more likely to be classified as low recidivists than blacks, even though they had reoffended.

---

[6] Part of the questionnaire is filled in by the defendants, which is why the developers responded to the related survey problem of "social desirability" by incorporating a "Lie Scale" in the form of supplementary questions.

The aforementioned study by Dressel and Farid (2018) found similar correlations with race, revealing that "Black" defendants were disproportionately more likely to receive false predictive values than "white" defendants. It is noteworthy that in accordance with anti-discrimination law, explicit information about race is not part of the COMPAS data set. However, as the two analyses have pointed out, there are many other indicators that seemingly have little to do with race (such as income, employment status, or prior convictions) but correlate highly with race and therefore act as proxies for race. Other studies found COMPAS to be less accurate for women and Hispanics, to the detriment of these groups (Hamilton 2019a, b). When statistical models are generated by software, race proxies may end up determining the model (Kirkpatrick 2017). In public debates, this insight was summarized by the statement that the discriminatory effects were not produced by the algorithm but by the data. The group-based exclusions and discriminations that emerge in societies—for example, owing to racial profiling in policing (see for example Silva and Deflem 2019), discrimination in education and the labor market, etc. (see for example Alexander and West 2012; Christin et al. 2015)—are inscribed in the data and then reproduced by software. Therefore, the seemingly neutral data about a defendant's employment, community, or marital status reflect the societal and organizational structures according to which he or she has been (partially) included in society.

3.2 Algorithmic Recommendations as Purposive Programs

Zooming in on the case of COMPAS, we see that the software *itself* works as a *conditional program*. Depending on the input information, a certain risk value is automatically assigned to defendants. But the de facto judicial *use* of the COMPAS score, as we argue, represents a more or less *purposive program*, at least formally: According to its developers (Northpointe 2015) and as outlined in a report about the introduction of COMPAS in various jurisdictions (Casey et al. 2014, p. 2), the score is not intended to replace but to "support" practitioners at crucial decision points within the criminal justice system, for instance, in placement decisions. Likewise, the Wisconsin Supreme Court has ruled (2016) that the COMPAS risk score *may* be considered in sentencing by judges, but it does not necessarily have to be taken into account. The score is therefore not a requirement but rather serves as one piece of information among others—in effect as an administratively introduced recommendation.

From the judges' perspective, we point out that the score is not only problematic with regard to statistical accuracy and "fairness" but above all because of the high level of uncertainty and ambiguity concerning the role that it should play in court decisions. Meaning-making requires information to be recognizable to others, and collectively acceptable explanations and accounts vary according to social context; hence, different institutions have different ways of reasoning that count as legitimate (Berger and Luckmann 1966; Mills 1940). In organizational contexts, it is essential for information to be uniformly "applicable" across different departments and situations (Tacke and Borchers 1993, p. 136). Collective sense-making of mediated information thus requires consistent and coherent organizational rules and

professional personnel that help to transform data into organizationally meaningful information. On these grounds, the COMPAS score is, first of all, organizationally "meaningless" because it is disconnected from the informational "richness" of the context from which it emerged. As a matter of fact, the mediatized reduction of communication in the form of a score, which is a highly aggregated form of data, is problematic, especially in organizations, because information does not simply have to be understood as such by the (legal) personnel (Henriksen and Bechmann 2020). In fact, the "human-in-the-loop" decision model embodies a paradox: Algorithms are developed to process an amount and complexity of data that surpass not only human agents but organizational boundaries, hence leaving the final decision to the very same decision-maker, who is, by definition, not in a position to meaningfully assess the data of the computer output.

Moreover, in the specific context of courts, and as mentioned above, sentencing decisions can be viewed as high-risk decisions about life and death that in particular take place under the condition of uncertainty. Consequently, these decisions are only made by judicial professionals and bound to individual case work and independent reasoning. In their decision-making, they have to apply the law by weighing various societal norms incorporated in the law when considering the differently provided facts in the claims. It is here that the legal profession shows that public administration is more than the logical execution and conditioned reading of law: To pursue the difficult task of subsumption, the physical co-presence of professional specialists and lay persons is key, as it allows for increased personal perceptions and cognition.

By asking direct questions—as well as by observing the language, facial expressions, and gestures throughout the procedural interaction of defendants, plaintiffs, and victims—judges seek to detect contradictions between data (text) and context and to add a whole background layer of meanings, inner attitudes, and memories. Machines could never do that. Thus, to apply the law according to professional standards, it is during the hearing in the courtroom that a judge is formally obliged to check the reliability of a defendant's presentation of self (including its comparison with those of the plaintiff and victims). Thus, interpreting perceptions is an extremely important function in organizations, especially when making discretionary and supplementary decisions in administrative contexts.

Against this backdrop, the COMPAS score can be regarded as decontextualized data that hardly help to absorb the uncertainty regarding both the plausibility and credibility of a defendant's narrative. On the contrary, as the questionnaire mostly reproduces (possibly discriminatory) data documented in previous parts of the file, it is not only negligible from a professional standpoint but also appears to be a redundant source of information (or even information overload) from the procedural and administrative perspective of the legal system, increasing judges' decisional complexity. Whereas Rudin et al. highlight that "the thought process of judges is (like COMPAS) a black box that provides inconsistent error-prone decisions" (2019, p. 25), it should be noted that court procedures are strongly formalized and that the judges' professional autonomy and neutrality are not only backed by the constitution but also covered by both relatively narrow organizational and procedural rules (namely the dominance of applying conditional programs instead of purposive programs).

To summarize, we have shifted the attention from the statistical production of scorings to their intra- and interorganizational reception. Our claim is that the communicative relevance of the COMPAS score for legal decision-making, in particular judicial subsumption, is not self-evident and the people using the score face the challenge of dealing with its many insecurities and ambiguities associated with their specific membership roles.

3.3 Patterns of Informal Behavior in Criminal Justice Organizations

The question remains whether and how the COMPAS score is factually granted relevance and authority in legal practice and its organizations. Few studies provide scientific evidence about the way in which the COMPAS score is actually integrated into judges' decisions. Among these, the ethnographic study by sociologist Angèle Christin (2017), in which she relies on observations, interviews, and document analysis, is particularly illuminating. As a core finding, Christin identifies an overall tendency of (reverse) "decoupling" (Meyer and Rowan 1977; see also Brunsson 1989) between "intended use and actual uses," which means in particular that judges minimize the impact of the COMPAS score in their daily work (Christin 2017, p. 8). Another recent empirical analysis confirms that the COMPAS score is perceived by court staff as a useful input, which is, however, routinely referred to as one among many informational elements when making a decision (Hartmann and Wenzelburger 2021). Christin observes three related strategies: foot-dragging, gaming, and open critique. In sociological research, such strategies can also be considered as "resistance through compliance" (Ybema and Horvers 2017): "While frontstage resistance derives its subversive potential from mixing open critique with implicit complaisance, backstage resistance functions via a benign appearance of carefully staged compliant behavior" (Ybema and Horvers 2017, p. 1233).

In what follows, we apply a theoretically inspired interpretation to Christin's case study. In doing so, we refine her main observations as an expression of the various intersections between "formal compliance" and "informal resistance" when using algorithmically generated information in organizations, ranging from open resistance to foot-dragging and criteria-tinkering. Similar to the result of Maiers' (2017) study, our analysis indicates that the users of data and analytics do not always make sense of and use analytics in the ways intended by their designers. We finally relate these practices to the phenomenon that scholars have termed the "social embeddedness of courts" (Ng and He 2017).

3.3.1 Open Resistance at the Outer Frontstage: Maintaining Professional Autonomy

Christin observes that in interviews, judges explicitly report that they ignore the risk score. They do so by referring to their judicial expertise, training, and experience (for a similar observation, see Angwin et al. 2016). Such claims can be understood as professional "accounts" (Garfinkel 1967; Orbuch 1997) that highlight the notion that judges are autonomous and not subjected to public opinion, political demands, or economic considerations. Here, another important difference between humans

and machines operating via decision premises becomes relevant—we will elaborate on this at the end of this section.

3.3.2 Foot-Dragging at the Inner Frontstage: Neutralizing the Relevance of Risk Scores

Christin (2017) also observes an informal refusal to consider COMPAS in criminal courts, which is a form of foot-dragging: She witnesses that even though risk scores were systematically added to defendants' files by pretrial and probation officers, none of the judges, prosecutors, or attorneys ever mentioned the scores during the hearings. Reading through the files of defendants, she found that the risk-score sheets were usually placed towards the end of the extensive files and, in contrast to the other pages of the file, were not annotated with hand-written remarks (Christin 2017, p. 9). This observation is consistent with those of other studies of decision-making processes based on algorithmic scores—such studies find that staff members engage in "conditioned reading" of scores, which they filter and temper in relation to other kinds of information, such as experience, intuition, and other pieces of "knowledge". In cases where the score is taken as a legitimate sign, it can be a powerful force in shaping the meaning of other signs of information and the categorization of defendants as being at a "high," "low," or "moderate" risk of recidivism. These conditioned readings have been observed in courts (Hartmann and Wenzelburger 2021) and hospitals (Maiers 2017).

3.3.3 Criteria-Tinkering at the Backstage: Synthetizing Tensions Within "Embedded Courts and Prisons"

Finally, judicial staff, more precisely probation officers, have developed strategies of "criteria-tinkering" (Hannah-Moffat et al. 2009, p. 405) by adapting the data that they enter into the risk-assessment tool in order to obtain the score that they thought was adequate for a given defendant (Christin 2017). This practice can be understood as an attempt to synthesize managerial requests, such as the score, and professional expectations of legal autonomy. Indeed, in many aspects, the COMPAS score reflects economic expectations in administration. Historically, the introduction of the COMPAS score was accompanied by managerial initiatives to (re-)gain an efficient allocation of resources: One of the aims was to reduce the number of incarcerations and assign convicted individuals who show little risk of re-offending to social programs. Economic rationales that take place at backstage of organized jurisdiction, including court management in general and the case flow in particular, can conflict with legal logics and lead judicial personnel to practices of symbolic compliance. In addition to criteria-tinkering, for instance, "legal professionals" sometimes redirect problematic files to alternative courts, subdivisions, or attorneys, in order to reduce their incarceration numbers and therefore "improve" their performance statistics (Christin 2017).

Comparing these three dynamics of formal and informal practices of (not) relating to the COMPAS score, an important difference between humans and machines operating via decision premises emerges: For lawyers, the process of drawing a con-

clusion from a fact and transforming it into a legal consequence is the final form in which they present their work results, but this is not a logic image or model of their factual decision-making activity. As legal norms are not ready-made conditional programs, the logical form has a representational function. In addition to the important function of perceiving, the actual legal decision-making performance thus consists in the interpretation of the decisional programs and the "incoming" information so that they comply with this requirement of representing (acceptable) decisions and risks to others. Put differently, legal decisions are less controlled by the process of their production and more by the requirements of their representation (Goffman 1959). This important fact is due to the special function of organized law in modern societies, as mentioned above. In contrast to private sector decisions, legal decisions take place in a political environment as courts produce binding decisions that have a direct external effect on each individual case and the subjects of that case—by law, these decisions may be neither improved nor worsened by using machines (Luhmann 1966, p. 20).

This structural compulsion to issue decisions and justifications gives rise to a strong disciplining of the (retro- and intersubjectively comprehensible) presenting of legal decision production that necessarily takes place as a means of collective uncertainty absorption. Uncertainty absorption occurs when information that comes from relatively uncertain decision premises is used as the basis for further decisions without recourse to the original information. This is precisely the specific function of the legal problem processing technique and its role protection, which allows decisions to be presented as correct (March and Simon 1958; see also March and Shapira 1987; Weick et al. 2005).

Legal uncertainty absorption also sheds light on one of the central requirements of modern organizations—to represent themselves in front of nonmembers. In this case, the function of automation resides in its symbolism (Espeland and Sauder 2007): To implement risk scores as purposive programs (and not as conditional programs) in courts may be seen as a facilitator for the intraorganizational connection of three different subsystems, the judicial, the political, and the managerial (or economic), with the result of competing logics and indissoluble tensions.[7] Against this background, the discourse on "algorithmic discrimination," with its focus on technological, statistical, and ethical aspects, entails the risk of neglecting or even undermining the struggles and trade-offs in organizations between economic, political, and professional rationalities.

---

[7] It is outside the scope of this article but noteworthy that algorithmic risk assessments are also used to evaluate jurisdiction itself. In the United States and the United Kingdom, it is common for judge's decisions to be analyzed in detail. In contrast, French law is aimed at preventing organizations from deploying litigation analytics and prediction, banning the publication of statistical information about French judges' decisions—with a 5-year prison sentence as the potential penalty in case of infringement (see ArtificialLawyer 2019).

4 Discussion and Conclusion

Our point of departure was the claim that the discourse about "algorithmic discrimination" leaves the fact widely unattended that software is embedded in organizations and shaped by formal and informal structures that both drive and limit the use of algorithmically generated data.

We point out that what is called "algorithmic discrimination" in the literature does not originate from algorithms as such, but either from behavioral data that reflects informal and particularistic norms in society and/or from the use of software that inscribes potentially discriminatory information (deriving from other social contexts that rely upon group-related distinctions) into other organizational decision structures. We therefore highlight that, when seeking to assess the risk of algorithms for discrimination, it is not enough to study the data and/or the mathematical model underlying the algorithms as such. It is equally important to identify the decision rules within an organization that exert (depersonalized) authority on algorithmically generated data.

To understand how group-based expectations take effect regarding the deployment of software in organizational contexts, we drew on organizational theory and introduced the distinction between conditional and purposive programs. Illustrating this distinction with the formal structures and informal practices of using the COMPAS score in US courts to assess the recidivism risk of defendants, we showed that the COMPAS score is not always a decisive factor but is often minimized and even ignored in legal decision-making owing to its organizational embeddedness as a purposive program. In contrast to conditional programs, purposive programs give decision-makers more leeway to consider different pieces of (algorithmically generated) information and therefore allow decision-makers to mitigate (or to aggravate) discrimination. Hence, adopting an organizational perspective allows scholars to challenge simple narratives, such as the hypothesis of a general over-reliance of human decision-makers on information generated by software.

These observations have numerous implications that mirror the main organizational boundaries of public administration: First, the question of information complexity and "errors" in partially automated settings that occur in contact with its "customers"; second, the problem of public accountability; and third, the questions of the economic efficiency of automation at the boundary to the political system.

First, when we study decision-making structures in organizations, we see that sophisticated predictive scores do not necessarily reduce information complexity and "errors" for decision-makers. In most cases, they may do both—they decrease and increase uncertainty—in particular when data analytics are introduced as a purposive program and thus used as one source among others. Decision-makers now have more (sometimes redundant) sources of information to consider, especially when there is more than one score at hand (Hartmann and Wenzelburger 2021), but they often lack sufficient capacity and guidance to select, interpret or question the often unduly ambiguous, redundant, and opaque status of behavioral data and of "algorithmic recommendations." Indeed, experimental evidence indicates that when faced with an increasing amount of information, human decision-makers make "less accurate" predictions about recidivism, because they are receiving, perceiving, and

processing "more" communication (Lin et al. 2020). Therefore, in the administrative systems of all kinds of support bodies, key staff members (police officers and welfare officers, judges, social workers, or supervisors) cooperate in short-information circuits to (re-)assess the usefulness of information for establishing the facts that make a decision program snap into place. This means, to avoid discrimination, the implementation of software in organizations should not only be subject to serious supervision—for example, by qualified data protection officers and other overseeing authorities—but possibilities for intervention and organizational learning should also be established (March 1991; see also Büchner and Dosdall 2021).

Second, societal requests for public accountability should be directed toward the structures and rules that condition that actual *use* of scores in organizations. Although most scoring systems are proprietary software and therefore not easily accessible for public scrutiny, the rules and practices employed when using a score are determined by the organizations in question. In the case of state agencies and courts, these organizations have an obligation to be responsive to public transparency requests. At the same time, formal rules cannot entirely eradicate unlawful practices, such as racial profiling in sentencing. Thus, as the algorithmic logics underlying technology and automation are inextricably interwoven with organizational structures, strategies to reduce algorithmic biases in data models will not necessarily trickle down to actual use in practice. Rather, this points us to the importance of adopting an organizational approach that investigates the role and dynamics of professions and personnel as "decisional premises" in organizations in follow-up studies—in addition to analyzing the purposive and conditional programs in organizations.

Third, our findings point toward a need for research that explores how the introduction of software into public administration is linked to market-oriented accounting frameworks—e.g., under the umbrella of "new public management" (Dunleavy et al. 2005; Katz 2012)—reflecting attempts to increase efficiency in the allocation of financial and human resource predictions. In this realm, it would be worth studying how the use of digital technologies (or even the discourse about it) camouflages previous organizational reforms (Brunsson and Jacobsson 2000; Brunsson and Olsen 1993) aimed at transforming rule-based bureaucracy to a performance-oriented, calculating "accounting entity." Various studies have shown that the incorporation of managerial practices into public administration may conflict with professional values; in the case of judges, the values in question would be the rehabilitation of convicts, the safety of victims, and societal security. In the same vein, it is necessary to explore the epistemic logics underlying data models and how conflicting organizational and professional norms (such as sanctioning versus social integration) are inscribed and weighed up in algorithms. Given that public administrations face pressures to become more efficient, to reduce their use of resources, and to legitimate their practices in accordance with (norms of) technological innovations, the implications we sketched may become more standardized. Professionals and scholars are thus more likely to (re)distribute scarce resources (attention, competencies, and time) on designing organizational contexts and decision premises that may leverage the advantages of data for nonlegal decisions while preserving the independence of legal subsumption and reasoning forms of knowledge as well.

Finally, in terms of further research, there is a need for more analyses that take into consideration the organizational uncertainty and ambivalence under which decisions are made and the structures within which they operate. Although many laboratory studies explore the conditions for human decision-making, there are very few field studies and ethnographies on the organizational development and use of software that take the intra- and interorganizational environments of decision-making, including efficiency pressures into account. Knowing that employees show very different degrees of awareness and critical distance toward algorithms and automation (Gran et al. 2020; Kolkman 2020), we also encourage analyses that compare how members of different professional occupations, in contrast to non- or semi-professionals, such as social workers or police officers, etc., deal with "algorithmic recommendations" in organizations. It is possible that members of established professions, such as judges and lawyers and doctors and nurses (see Maiers 2017), show relatively high autonomy vis-à-vis machine recommendations when taking decisions, and coping with "external" (political, administrative, or managerial) norms.

To conclude, organization matters with regard to the discriminatory effects of algorithms. The data and the calculative models are far from the only components of what is misleadingly understood as algorithm-driven decisions. The formal and informal structures of the organization, including the type of decision premises and decision programs, as well as the societal organizational context and the professional identities, need to be considered when searching for ways to mitigate discrimination.

References

Abbott, Andrew. 1988. *The system of professions: an essay on the division of expert labor*. Chicago: University of Chicago Press.

Ackroyd, Stephen. 2016. Sociological and organisational theories of professions and professionalism. In *The Routledge companion to the professions and professionalism*, ed. Mike Dent, Ivy Lynn Bourgeault, Jean-Louis Denis and Ellen Kuhlmann, 15–30. London: Routledge.

Alaimo, Cristina, and Jannis Kallinikos. 2020. Managing by data: algorithmic categories and organizing. *Organization Studies* https://doi.org/10.1177/0170840620934062.

Alexander, Michelle, and Cornel West. 2012. *The new Jim Crow: mass incarceration in the age of colorblindness*. New York: New Press.

Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner. 2016. Machine bias. ProPublica, May 23. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (Accessed 1 June 2021).

Angwin, Julia, Noam Schreiber and Ariana Tobin. 2017. Dozens of companies are using Facebook to exclude older workers from job ads. ProPublica, December 20. https://www.propublica.org/article/facebook-ads-age-discrimination-targeting (Accessed 1 June 2021).

Araújo, Camila Souza, Meira Wagner Jr. and Virgilio Almeida. 2016. Identifying stereotypes in the online perception of physical attractiveness. In *Social Informatics*, Vol. 10046, eds. Emma Spiro and Yong-Yeol Ahn, 419–437. Cham: Springer International Publishing.

ArtificialLawyer. 2019. France bans judge analytics, 5 years in prison for rule breakers. January 1. https://www.artificiallawyer.com/2019/06/04/france-bans-judge-analytics-5-years-in-prison-for-rule-breakers.

Avery, Robert B., Kenneth P. Brevoort and Glenn Canner. 2012. Does credit scoring produce a disparate impact? *Real Estate Economics* 40:65–114.

Beckert, Jens, and Patrik Aspers (eds.). 2011. *The worth of goods: valuation and pricing in the economy*. New York: Oxford University Press.

Benjamin, Ruha. 2019. *Race after technology: abolitionist tools for the new Jim Code*. Medford: Polity.

Berger, Peter L., and Thomas Luckmann. 1966. *The social construction of reality. A treatise in the sociology of knowledge*. New York: Doubleday.

Beverungen, Armin, Timon Beyes and Lisa Conrad. 2019. The organizational powers of (digital) media. *Organization* 26:621–635.

Brunsson, Nils. 1989. *The organization of hypocrisy: talk, decisions, and actions in organizations*. Chichester: John Wiley.

Brunsson, Nils, and Bengt Jacobsson (eds.). 2000. *A world of standards*. Oxford: Oxford University Press.

Brunsson, Nils, and Johan P. Olsen. 1993. *The reforming organization: making sense of administrative change*. London: Routledge.

Büchner, Stefanie. 2018. Digitale Infrastrukturen – Spezifik, Relationalität und die Paradoxien von Wandel und Kontrolle. *Arbeits- und Industriesoziologische Studien* 11:279–293.

Büchner, Stefanie, and Henrik Dosdall. 2021. Organisation und Algorithmus: Wie algorithmische Kategorien, Vergleiche und Bewertungen durch Organisationen relevant gemacht werden. In *Soziale Praktiken des Beobachtens: Vergleichen, Bewerten, Kategorisieren und Quantifizieren* Sonderheft 61, Kölner Zeitschrift für Soziologie und Sozialpsychologie, ed. Bettina Heintz and Theresa Wobbe, 333–357. Wiesbaden: Springer VS.

Caliskan, Aylin, Joanna J. Bryson and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356:183–186.

Caplan, Robyn, and dana boyd. 2018. Isomorphism through algorithms: institutional dependencies in the case of Facebook. *Big Data & Society* 5:1–12.

Carruthers, Bruce G., and Laura Ariovich. 2010. *Money and credit: a sociological approach*. Cambridge: Polity.

Casey, Pamela M., Jennifer K. Elek, Roger K. Warren, Fred Cheesman, Matt Kleiman and Brian Ostrom. 2014. *Offender risk & needs assessment instruments: a primer for courts*. Williamsburg: National Center for State Courts. https://www.ncsc.org/__data/assets/pdf_file/0018/26226/bja-rna-final-report_combined-files-8-22-14.pdf.

Cech, Florian, Fabian Fischer, Soheil Human, Paola Lopez and Ben Wagner. 2019. Dem AMS-Algorithmus fehlt der Beipackzettel. Futurezone, October 3. https://futurezone.at/meinung/dem-ams-algorithmus-fehlt-der-beipackzettel/400636022 (Accessed 1 June 2021).

Chouldechova, Alexandra. 2017. Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. *Big Data* 5:153–163.

Christin, Angèle. 2017. Algorithms in practice: comparing web journalism and criminal justice. *Big Data & Society* 4:1–14.

Christin, Angèle, Alex Rosenblat and dana boyd. 2015. Courts and predictive algorithms. *Data & Civil-Right* 1–13.

Cohen, Michael D., G. March James and Johan P. Olsen. 1972. A garbage can model of organizational choice. *Administrative Science Quarterly* 17(1):1–25.

Cohen, Michael D., G. March James and Johan P. Olsen. 1994. Ambiguity and choice in organizations. In *Ambiguity and choice in organizations*, ed. James G. March, Johan P. Olsen, 24–37. Oslo: University Press.

Couldry, Nick, and Jun Yu. 2018. Deconstructing datafication's Brave New World. *New Media & Society* 20:4473–4491.

Daston, Lorraine, and Peter P. Galison. 2007. *Objectivity*. New York: Zone Books.

DiMaggio, Paul J., and Walter W. Powell. 1983. The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review* 48:147–160.

Dolata, Ulrich. 2019. Privatization, curation, commodification: commercial platforms on the Internet. *Österreichische Zeitschrift für Soziologie* 44:181–197.

Dourish, Paul. 2016. Algorithms and their others: algorithmic culture in context. *Big Data & Society* 3:1–11.

Dressel, Julia, and Hany Farid. 2018. The accuracy, fairness, and limits of predicting recidivism. *Science Advances* 4:1–5.

Dunleavy, Patrick, Helen Margetts, Simon Bastow and Jane Tinkler. 2005. New public management is dead. Long live digital-era governance. *Journal of Public Administration Research and Theory* 16:467–494.

Edelman, Benjamin, Michael Luca and Dan Svirsky. 2017. Racial discrimination in the sharing economy: evidence from a field experiment. *American Economic Journal: Applied Economics* 9:1–22.

Espeland, Wendy N., and Michael Sauder. 2007. Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology* 113:11–40.

Etzioni, Amitai. 1961. *A comparative analysis of complex organizations: on power, involvement, and their correlates*. New York: Free Press.

Etzioni, Amitai. 1969. *The semi-professions and their organization: teachers, nurses, social workers*. New York: Free Press.

Feldman, Martha S., and James G. March. 1981. Information in organizations as signal and symbol. *Administrative Science Quarterly* 26:171–186.

Foot, Kirsten A., Pablo J. Boczkowski and Tarleton Gillespie (eds.). 2014. *Media technologies: essays on communication, materiality, and society*. Cambridge: MIT Press.

Funder, Maria. 2018. Neo-Institutionalismus und geschlechterorientierte Organisationsforschung – Befunde und Plädoyer für einen weiterführenden Dialog. In *Geschlecht als widersprüchliche Institution: Neo-institutionalistische Implikationen zum Gender Cage in Organisationen*, eds. Nathalie Amstutz, Helga Eberherr, Maria Funder and Roswitha Hofmann, 305–342. Baden-Baden: Nomos.

Garfinkel, Harold. 1967. *Studies in ethnomethodology*. Cambridge: Polity.

Gigerenzer, Gerd, and Christoph Engel (eds.). 2006. *Heuristics and the law*. Cambridge: MIT Press.

Gilheany, John, David Wang and Stephen Xi. 2015. The model minority? Not on Airbnb.com: A hedonic pricing model to quantify racial bias against Asian Americans. *Technology Science* 1–19.

Gillespie, Tarleton. 2014. The relevance of algorithms. In *Media technologies: essays on communication, materiality, and society*, eds. Kirsten A. Foot, Pablo J. Boczkowski and Tarleton Gillespie, 167–193. Cambridge: MIT Press.

Goffman, Erving. 1959. *The presentation of self in everyday life*. New York: Anchor Books.

Graeber, David. 2016. *The utopia of rules: on technology, stupidity, and the secret joys of bureaucracy*. Brooklyn: Melville House.

Gran, Anne-Britt, Peter Booth and Taina Bucher. 2020. To be or not to be algorithm aware: A question of a new digital divide? *Information, Communication & Society* 1–18.

Hamilton, Melissa. 2019a. The biased algorithm: evidence of disparate impact on hispanics. *The American Criminal Law Review* 56:1553.

Hamilton, Melissa. 2019b. The sexist algorithm. *Behavioral Sciences & The Law* 37:145–157.

Hannah-Moffat, Kelly, Paula Maurutto and Sarah Turnbull. 2009. Negotiated risk: actuarial illusions and discretion in probation. *Canadian Journal of Law and Society* 24:391–409.

Hartmann, Kathrin, and Georg Wenzelburger. 2021. Uncertainty, risk and the use of algorithms in policy decisions: a case study on criminal justice in the USA. *Policy Sciences* 54:269–287.

Heintz, Bettina. 2008. Governance by Numbers: Zum Zusammenhang von Quantifizierung und Globalisierung am Beispiel der Hochschulpolitik. In *Governance von und durch Wissen*, ed. Gunnar Folke Schuppert, Andreas Voßkuhle, 110–129. Baden-Baden: Nomos.

Heintz, Bettina, and Theresa Wobbe (eds.). 2021. *Soziale Praktiken des Beobachtens: Vergleichen, Bewerten, Kategorisieren und Quantifizieren*. Sonderheft 61, Kölner Zeitschrift für Soziologie und Sozialpsychologie. Wiesbaden: Springer VS.

Henriksen, Anne, and Anja Bechmann. 2020. Building truths in AI: making predictive algorithms doable in healthcare. *Information, Communication & Society* 23:802–816.

Hofstra, Bas, and Niek C. Schipper. 2018. Predicting ethnicity with first names in online social media networks. *Big Data & Society* 5:1–14.

Katz, Daniel Martin. 2012. Quantitative legal prediction—or—how I learned to stop worrying and start preparing for the data driven future of the legal services industry. *Emory Law Journal* 62:909–966.

Katzenbach, Christian, and Lena Ulbricht. 2019. Algorithmic governance. *Internet Policy Review* 8:1–18. Special issue defining concepts of the digital society.

Kay, Matthew, Cynthia Matuszek and Sean A. Munson. 2015. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference*

on Human Factors in Computing Systems, ed. Bo Begole, Jinwoo Kim, Kori Inkpen and Woontack Woo, 3819–3828.

Kehl, Danielle, Priscilla Guo and Samuel Kessler. 2017. *Algorithms in the criminal justice system: assessing the use of risk assessments in sentencing*. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.

Kette, Sven. 2008. *Bankenregulierung als Cognitive Governance: Eine Studie zur gesellschaftlichen Verarbeitung von Komplexität und Nichtwissen*. Wiesbaden: Springer VS.

Kirkpatrick, Keith. 2017. It's not the algorithm, it's the data. *Communications of the ACM* 60:21–23.

Kolkman, Daan. 2020. The (in)credibility of algorithmic models to non-experts. *Information, Communication & Society* 1–17.

Lambrecht, Anja, and Catherine Tucker. 2016. *Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of STEM career ads*. Social Science Research Network. https://doi.org/10.2139/ssrn.2852260.

Lin, Zhiyuan, Jung Jongbin, Sharad Goel and Jennifer Skeem. 2020. The limits of human predictions of recidivism. *Science Advances* 6:1–8.

Lowry, Stella, and Gordon MacPherson. 1988. A blot on the profession. *British Medical Journal* 296:657–658.

Luhmann, Niklas. 1966. *Recht und Automation in der öffentlichen Verwaltung: eine verwaltungswissenschaftliche Untersuchung*. Berlin: Duncker & Humblot.

Luhmann, Niklas. 1983. *Legitimation durch Verfahren*. Frankfurt: Suhrkamp.

Luhmann, Niklas. 2018. *Organization and decision*. Cambridge: Cambridge University Press.

Maiers, Claire. 2017. Analytics in action: users and predictive data in the neonatal intensive care unit. *Information, Communication & Society* 6:1–16.

March, James G. 1991. Exploration and exploitation in organizational learning. *Organization Science* 2:71–87.

March, James G., and Zur Shapira. 1987. Managerial perspectives on risk and risk taking. *Management Science* 33:1404–1418.

March, James G., and Herbert A. Simon. 1958. *Organizations*. New York: Wiley.

Mehrpouya, Afshin, and Rita Samiolo. 2016. Performance measurement in global governance: ranking and the politics of variability. *Accounting, Organizations and Society* 55:12–31.

Mennicken, Andrea. 2013. 'Too big to fail and too big to succeed': accounting and privatisation in the prison service of England and Wales. *Financial Accountability & Management* 29:206–226.

Meyer, John W., and Brian Rowan. 1977. Institutionalized organizations: formal structure as myth and ceremony. *American Journal of Sociology* 83:340.

Mills, C. Wright. 1940. Situated actions and vocabularies of motive. *American Sociological Review* 5:904.

Mitchell, Shira, Eric Potash, Solon Barocas, Alexander D'Amour and Kristian Lum. 2021. Algorithmic fairness: choices, assumptions, and definitions. *Annual Review of Statistics and its Application* 8:141–163.

Monahan, John, and Jennifer L. Skeem. 2016. Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology* 12:489–513.

Ng, Kwai Hang, and Xin He. 2017. *Embedded courts: judicial decision-making in China*. Cambridge: Cambridge University Press.

Noble, Safiya Umoja. 2018. *Algorithms of oppression: how search engines reinforce racism*. New York: New York University Press.

Northpointe. 2015. Practitioner's guide to COMPAS core. http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf.

Orbuch, Terri L. 1997. People's accounts count: the sociology of accounts. *Annual Review of Sociology* 23:455–478.

Perrow, Charles. 1991. A society of organizations. *Theory and Society* 20:725–762.

Poon, Martha. 2007. Scorecards as devices for consumer credit: the case of Fair, Isaac & Company Incorporated. *The Sociological Review* 55:284–306.

Rona-Tas, Akos, and Stefanie Hiss. 2010. The role of ratings in the subprime mortgage crisis: The art of corporate and the science of consumer credit rating. In *Markets on Trial: The Economic Sociology of the U.S. Financial Crisis: Part A* Research in the Sociology of Organizations, Vol. 30 Part A, ed. Michael Lounsbury, Paul M. Hirsch, 115–155. Bingley: Emerald.

Rosenblat, Alex, Tamara Kneese and dana boyd. 2014. Networked employment discrimination. Open Society Foundations' future of work commissioned research papers. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2543507.

Rosenblat, Alex, Karen E.C. Levy, Solon Barocas and Tim Hwang. 2017. Discriminating tastes: Uber's customer ratings as vehicles for workplace discrimination: customer ratings and workplace discrimination. *Policy & Internet* 9:256–279.

Rudin, Cynthia, Caroline Wang and Beau Coker. 2019. The age of secrecy and unfairness in recidivism prediction. *Harvard Data Science Review* 2:1–46.

Schwarting, Rena. 2020. *Organisationsbildung und gesellschaftliche Differenzierung: Empirische Einsichten und theoretische Perspektiven*. Wiesbaden: Springer VS.

Schwarting, Rena. 2010. *Risk transformation in financial organizations. A functional comparison of third-party observations of creditworthiness using credit rating grades and CDS prices*. University of Bielefeld. Faculty of Sociology.

Silva, Derek M.D., and Mathieu Deflem. 2019. Exporting preemption: the transnational diffusion of counter-radicalisation policing strategies. In *The development of transnational policing: past, present and future*, eds. John L.M. McDaniel, Karlie E. Stonard and David J. Cox, 179–200. New York: Routledge.

Simon, Herbert A. 1947. *Administrative behavior: a study of decision-making processes in administrative organizations*. New York: Free Press.

Simon, Herbert A. 1991. Bounded rationality and organizational learning. *Organization Science* 2:125–134.

Sinclair, Timothy. 2008. *The new masters of capital: American bond rating agencies and the politics of creditworthiness*. Ithaca: Cornell University Press.

Strulik, Torsten, and Helmut Willke. 2007. *Towards a cognitive mode in global finance: the governance of a knowledge-based financial system*. Frankfurt: Campus.

Supreme Court of Wisconsin. 2016. *2015AP157-CR: State of Wisconsin v. Eric L. Loomis*

Sweeney, Latanya. 2013. Discrimination in online ad delivery. *Communications of the Association of Computing Machinery* 56:44–54.

Tacke, Veronika. 2008. Neutralisierung, Aktualisierung, Invisibilisierung: Zur Relevanz von Geschlecht in Systemen und Netzwerken. In *Geschlechterdifferenzen – Geschlechterdifferenzierungen. Ein Überblick über gesellschaftliche Entwicklungen und theoretische Positionen*, ed. Sylvia Marlene Wilz, 253–289. Wiesbaden: VS.

Tacke, Veronika, and Uwe Borchers. 1993. Organisation, Informatisierung und Risiko: Blinde Flecken mediatisierter und formalisierter Informationsprozesse. In *Risiken informatisierter Produktion, Beiträge zur sozialwissenschaftlichen Forschung*, ed. Hans-Jürgen Weißbach, Andrea Poy, 125–151. Wiesbaden: VS.

Thompson, James D. 1967. *Organizations in action: social science bases of administrative theory*. New Brunswick: Transaction.

Weick, Karl E., Kathleen M. Sutcliffe and David Obstfeld. 2005. Organizing and the process of sensemaking. *Organization Science* 16:409–421.

Ybema, Sierk, and Martha Horvers. 2017. Resistance through compliance: the strategic and subversive potential of frontstage and backstage resistance. *Organization Studies* 38:1233–1251.

Yeung, Karen. 2018. Algorithmic regulation: a critical interrogation. *Regulation & Governance* 12:505–523.

Zald, Mayer N. 1990. History, sociology and theories of organization. In *Institutions in American society: essays in market, political, and social organizations*, ed. John Edgar Jackson, 81–108. Ann Arbor: University of Michigan Press.

Rena Schwarting Dr., ist Organisationssoziologin und International Fellow an der Universität St. Gallen (Schweiz) sowie Gastwissenschaftlerin in der Forschungsgruppe "Politik der Digitalisierung" am Wissenschaftszentrum Berlin für Sozialforschung (WZB). Schwerpunkte ihrer Forschungen sind Organisationsbildungen und der Einsatz von digitalen Technologien in und durch Organisationen. Ausgewählte Publikationen: Organisationsbildung und gesellschaftliche Differenzierung. Wiesbaden 2020; Organisation und Verfahren. Zur Legitimation von Entscheidungen zwischen Interaktion und Gesellschaft. Soziale Systeme 22 (Sonderheft), 2017; Hochfrequenzhandel zwischen Entscheidungsautomation und Entscheidungsautonomie. In: Organisation und Unsicherheit, Hrsg. M. Apelt und K. Senge. Wiesbaden 2014.

Lena Ulbricht Dr., ist Politikwissenschaftlerin und promovierte an der Humboldt-Universität zu Berlin. Ihr Habilitationsprojekt beschäftigt sich mit dem Einsatz künstlicher Intelligenz für staatliche Regulierung und seinen demokratischen Implikationen. Ihre Forschungsinteressen sind Regulierung und Governance, Public Policy, vergleichende Forschung, Big Data und künstliche Intelligenz, Datenschutz, kritische Sicherheitsstudien, Wissenschaftssoziologie, Bildungsforschung und Forschung über Sozialpolitik. Aktuelle Publikationen: The Weizenbaum Journal of the Digital Society: Editorial. The Weizenbaum Journal of the Digital Society 1, 2021 (mit M. Emmer, H. Krasnova, M. Krzywdzinski, A. Metzger, S. Schimmler und C. Neuberger); Algorithmic regulation: A maturing concept for investigating regulation of and through algorithms. Regulation & Governance, Special Issue "Algorithmic Regulation" 2021 (als Hrsg. mit K. Yeung).