Transparency as Manipulation? Uncovering the Disciplinary Power of Algorithmic Transparency

Hao Wang

Received: 15 November 2021 / Accepted: 12 July 2022 © The Author(s) 2022

Abstract
Automated algorithms are silently making crucial decisions about our lives, but most of the time we have little understanding of how they work. To counter this hidden influence, there have been increasing calls for algorithmic transparency. Much ink has been spilled over the informational account of algorithmic transparency—about how much information should be revealed about the inner workings of an algorithm. But few studies question the power structure beneath the informational disclosure of the algorithm. As a result, the information disclosure itself can be a means of manipulation used by a group of people to advance their own interests. Instead of concentrating on information disclosure, this paper examines algorithmic transparency from the perspective of power, explaining how algorithmic transparency under a disciplinary power structure can be a technique of normalizing people’s behavior. The informational disclosure of an algorithm can not only set up some de facto norms, but also build a scientific narrative of its algorithm to justify those norms. In doing so, people would be internally motivated to follow those norms with less critical analysis. This article suggests that we should not simply open the black box of an algorithm without challenging the existing power relations.

Keywords Algorithm · Transparency · Manipulation · Disciplinary power · Surveillance

1 Introduction
Automated algorithms have been pervasive and consequential in determining our lives, from loan approvals and job applications to legal decisions and college admissions (Broussard, 2020; Calo & Citron, 2021; Citron & Pasquale, 2014). But most of the time we have a very limited understanding of how the algorithms work (Burrell,
These black-boxed algorithms can not only hide the problems of inaccuracy and discrimination in the decision-making process (Citron & Pasquale, 2014; Zarsky, 2016; Binns, 2018; Eubanks, 2018; Franke, 2022), but they may also manipulate individuals in both commercial and political fields (Susser et al., 2019; Zuboff, 2019; Sax, 2021). In response, it seems natural to call for algorithmic transparency in order to counter the pernicious influence that algorithms appear to have on decisions we make, or decisions that are made about us (Jauernig et al., 2022; von Eschenbach, 2021). Algorithmic transparency has been so much acclaimed in the circles of academics, professionals, regulators, and activists that to make algorithms transparent seems to be a universal good (Estop, 2014; Weller, 2017; Springer & Whittaker, 2019; Zerilli et al., 2019).

However, many scholars have challenged this standpoint, considering some practical issues about making algorithms transparent (Matzner, 2017; Ananny & Crawford, 2018; De Laat, 2018; Kemper and Kolkman, 2019; Diakopoulos, 2020; Powell, 2021; Kim & Moon, 2021). For example, some data about algorithms may be too sensitive to be disclosed (De Laat, 2018). The disclosure of companies’ algorithms may also hurt their competitive edge (Citron & Pasquale, 2014). Moreover, to make algorithm transparent may lead to consumers gaming the system (Diakopoulos, 2020). Although these objections are significant, they only tell part of the story. The notion of algorithmic transparency that they use is only seen as an information disclosure, concerning how much information an algorithm should be revealed. But what if algorithmic transparency itself is a form of manipulation? What if some algorithm deployers intentionally disclose the workings of their algorithms in order to operate its power of control and to only meet their own interests? In fact, algorithmic transparency is a social process intertwined with power relations, which is much more complex than the solely informational disclosure (Ananny & Crawford, 2018; Weller, 2017).

Instead of focusing on information disclosure, this paper will examine algorithmic transparency from the perspective of power. I will argue that algorithmic transparency under a disciplinary power structure can be a technique of normalizing people’s behavior. The argument focuses on a specific case, namely the FICO Score, the most widely used credit scoring tool in the USA. Unlike what is commonly understood, FICO’s scoring algorithm is not completely black-boxed (Citron & Pasquale, 2014). On the contrary, the company actively reveals the workings of its algorithms with a plethora of details to de-mystify its scoring process. The paper tries to show that it is precisely through this apparent act of transparency that the disciplinary power of FICO Score effectively works.

The paper mainly builds on a case study of credit surveillance in the USA, but it reflects a general phenomenon that algorithmic transparency can be a disciplinary practice.

1 In this paper, I follow the understanding of manipulation given by Daniel Susser, Beate Roessler, and Helen Nissenbaum: “manipulation functions by exploiting the manipulee’s cognitive (or affective) weaknesses and vulnerabilities in order to steer his or her decision-making process towards the manipulator’s ends.” (Susser, Roessler & Nissenbaum, 2019, 3).
technique in the era of Big Data. For Danielle Citron and Frank Pasquale, credit scoring system is a typical example of what they called the “Scored Society” (Citron & Pasquale, 2014). Citron and Pasquale show that there is a “scoring trend” in Big Data society, where predictive algorithms are used “to rank and rate individuals” in countless areas of life (Citron & Pasquale, 2014, 1). Various scoring algorithms are developed to rank which candidates are the perfect fit for a job position (Black & van Esch, 2020; Kim, 2020), how likely individuals would commit crimes (McKay, 2020; Schwerzmann, 2021) or vote for a candidate (Susser et al., 2019), and even how likely potential daters are matched (Tuffley, 2021).

These scoring systems are often embedded with a series of punishments, which may discipline individuals’ behavior. If some people are rated as high risk of committing crimes, they may be arrested or not be allowed to fly on a plane (Lyon, 2007). If individuals are algorithmically ranked as health risks, they would be soon charged higher insurance premiums (Citron & Pasquale, 2014, 4). If Uber drivers have relatively low scores, they can be immediately punished by the algorithm-driven platform which can restrict the numbers of passengers they receive, or even eventually remove the drivers from the platform (Chan, 2019; O’Connor, 2022; Muldoon & Raekstad, 2022). With these punishments, not only are these scoring systems used for the powerful entities to make relevant decisions about individuals, but the individuals themselves are also subject to disciplinary process by means of algorithmic transparency. Under the disciplinary structure of Uber, for instance, the disclosure of Uber’s scoring algorithm may make drivers know what the rules are to maximize their ratings, so they can be more easily subject to the disciplinary power.

It is noteworthy that the paper’s concentrating on the particular case of the FICO Score does not mean that only commercial entities can implement manipulation through algorithmic transparency. Governments can also use algorithmic transparency to manipulate its citizens, a typical example of which is China’s Social Credit System (Creemers, 2018; Dai, 2018; Ding and Zhong, 2021; Zou, 2021; Drinhausen & Brussee, 2021). For instance, the Osmanthus Score is a local version of social credit system conducted by the Chinese local authority in Suzhou, a southern-east city near Shanghai. Not only does the Osmanthus Score explain what range of data its algorithm calculates, but it also illustrates clearly how its score can be built up “for donating blood, volunteer work, and winning award or special honors” and can be lowered down if citizens are blacklisted by any of the central or regional authorities (Ahmed, 2018, 50). By disclosing such information, citizens would not simply passively wait to be evaluated, but they can know what the state expects them to do to improve their scores, so they would more actively participate in the disciplinary process. Unlike companies that often discipline consumers’ behavior toward their most profitable outcomes, governments may tend to operate such disciplinary power for overall governance of their citizens’ behavior.

This article consists of four parts. It starts with a critical analysis of the common use of “algorithmic transparency” from an account of power. Part two introduces the case of FICO’s algorithmic transparency in terms of information disclosure and shows why such transparency is problematic. Part three demonstrates that the FICO Score is fundamentally a disciplinary system, uncovering the disciplinary power beneath the surface of information disclosure. Part four shows how FICO’s
algorithmic transparency functions as a disciplinary technique to normalize individuals to be responsible credit consumers. The paper suggests that we should not just turn to an informational disclosure of algorithms without challenging the existing power structure; otherwise, algorithmic transparency itself may turn out to be a technique of control.

2 Two Accounts of Algorithmic Transparency

In computer science, an algorithm basically follows the “if…then…else” logic: “If a happens, then do b; if not, then do c” (Bucher, 2018; Smith, 2018). In the last decade, the meaning of algorithm has become associated with complex decision-making process used by various automated machines (Baum et al., 2022; Erasmus et al., 2021; O’Neil, 2016; Završnik, 2021). But an algorithm can be roughly viewed as “an ordered set of steps followed in order to solve a particular problem or to accomplish a defined outcome” (Diakopoulos, 2018, 2). The term “algorithmic transparency” then is “the disclosure of information about algorithms to enable monitoring, checking, criticism, or intervention by interested parties” (Diakopoulos & Koliska, 2016, 3).

This common understanding of algorithmic transparency assumes an informational account of transparency. Such an account centers on the role of information, describing transparency as a matter of information sharing and information disclosure. For instance, in Diakopoulos’s words, “transparency is merely about producing information” (Diakopoulos, 2020, 212). It offers “the informational substrate for ethical deliberation of a system’s behavior by external actors” (Diakopoulos, 2020, 198). To make an automated system transparent, for example, the information about the inner workings of the algorithm should be provided to relevant stakeholders. In this regard, the prime concern about transparency is the quantity and quality of the disclosed information. For example, Daniel Susser proposes a “radical transparency” to allow affected parties “a view under the hood” in order to disclose all relevant information (Susser, 2019, 5). Paul de Laat proposes a “limited transparency” that algorithm deployers only “open up their models to the public and provide reasons for decisions upon request” (De Laat, 2019, 9).

These scholars may have different points on how much of the internal workings of an algorithm should be disclosed, but they tend to hold that transparency is only a matter of information disclosure. This view of transparency as information disclosure has long been criticized (Prat, 2005; Drucker & Gumpert, 2007; Etzioni, 2010; Meijer, 2013; Estop, 2014; Flyverbom et al., 2015; Albu & Flyverbom, 2016; Weller, 2017; Ananny & Crawford, 2018). In such a view, transparency is deemed as somewhat a linear transmission of information where “a sender crafts a message or a collection of data and transmits it via a given channel to a receiver” (Albu & Flyverbom, 2016, 6). It presumes an ideal process of information transmission where the channel is stable, the sender’s information is accurate, and the receiver is capable to critically understand the provided information. However, this assumption of an ideal transmission is untrue. It fails to capture the complex phenomenon of transparency as a social process, which is not just a “precise end state in which everything is clear and apparent,” but also an operation of power (Ananny & Crawford, 2018, 3).
Power can operate through transparency to manipulate people—not only through hidden lies, but through the transparency of “truth” (Estop, 2014). For example, Adrian Weller warns that companies can manipulate users to trust its system by providing “an empty explanation as a psychological tool to soothe users” (Weller, 2017, 57). He also notes that strategic transparency can lead users into certain actions or behaviors. For instance, “Amazon might recommend a product, providing an explanation in order that you will then click through to make a purchase” (Weller, 2017, 56). A further concern is that transparency can reinforce existing power structures, instead of making them more accountable. When the dominant structure is dogmatic, it is not only useless to promote transparency, but can re-strengthen the existing power asymmetry (Kossow et al., 2021). As David Heald shows, if corruption in society goes on after being transparent, “knowledge from greater transparency may lead to more cynicism,” which may lead to wider corruption (Heald, 2006, 35). In this sense, transparency can reveal power asymmetries, but if the unbalanced power structure is deep-rooted and hard to change, then more transparency may in turn reinforce the existing power.

Nonetheless, this critical understanding of transparency, especially from the particular perspective of power, has not been much investigated particularly on the issues derived from algorithms and Big Data analytics. The main reason may be that algorithms are often seen as scientific and technical facts, which are relatively neutral and isolated from social and political power. However, as David Beer rightly points out, it is a “mistake” to see the algorithm solely as “technical and self-contained object” (Beer, 2017, 4). An algorithm, more than lines of code and data, is “an assemblage of human and non-human actors” (Ananny & Crawford, 2018, 11). The data that mined in the application of an algorithm is already biased by social judgments, which reflect the social world driven by interests and power (Matzner, 2017). When applied, algorithms are used by different companies with various intentions, to make profits, to control, or to manipulate, etc. So, to make algorithms transparent is not just about revealing objective information about how it works, but also about the interests of those who created it and their views about those who are to be subject to it.

Therefore, I consider it to be important to propose a particular power account on algorithmic transparency in this paper. Weller defines three classes of people typically involved in the process of algorithmic disclosure. The first group includes the developers who build the algorithms. The second group consists of deployers who own the algorithms and release them to the public or users. The third is users who are often affected by the algorithmic decisions. He uses the example of Amazon, in which “developers might be hired to build a personalized recommendation system to buy products, which Amazon then deploys, to be used by a typical member of the public” (Weller, 2017, 55). He notes that the intended audience (usually the users) for algorithmic transparency is not always the beneficiary (usually the deployers). So, there is a room for deployers to manipulate users through disclosing their algorithms. Weller reminds us that, “Even if a faithful explanation is given, it may have been carefully selected from a large set of possible faithful explanations in order to serve the deployer’s goals” (Weller, 2017, 57).
In sum, algorithmic transparency is not merely about the informational disclosure of the inner workings of an algorithm. Instead, it also means *how power operates in the process of making the information disclosed*. Notedly, such a power analysis of algorithmic transparency does not mean that it is superior to the informational account or it can fully replace the latter. Rather, these two accounts are complementary, and both can be useful in illustrating different issues. The upshot is that when analyzing algorithmic transparency, we should take both accounts into consideration. We should not only disclose the information about how algorithms work, but also be alert to the hidden power structures and the way in which the disclosure happens can have profound and far-reaching effects that are often overlooked. Using the example of the FICO’s credit scoring algorithm, I will argue that algorithmic transparency itself needs to be critically analyzed to uncover the power relations which shape our lives, both in the online and offline worlds.

3 The Informational Disclosure of FICO’s Algorithm

When individuals apply for a credit card, car loan, mortgage, or other types of credit, lenders need to measure the risk of loaning them the money. In the USA, over 90% of lenders use FICO Scores to evaluate this risk, to decide:

a) Whether to approve you and;

b) What terms and interest rates you qualify for.\(^2\)

A FICO Score looks simple: it is a three-digit number, ranging from 300 to 850, which is meant to express the creditworthiness of the individual. A high credit score (i.e., above 800) is meant to indicate that a credit applicant is at low risk of default, while a low number indicates a higher risk. However, the inner working of a FICO Score is “based on complex, scientific algorithmic assessment” of the mined information.\(^3\)

Like other algorithmic decision-making systems, the FICO Score is often seen as a type of black box: the algorithm processes inputs and gives the output in the form of scores, without offering the user an insight into the logic of that algorithm (Pasquale, 2015). In reality, however, the scoring algorithm of FICO Score is not completely black-boxed. The FICO website clearly announces that it will de-mystify its scoring process for the benefit of the user: “credit scores don’t have to be a mystery—myFICO is here to help you to learn and control it.”\(^4\) It is not so much a statement as a bluff since FICO does disclose many meaningful explanations about how its algorithm works. FICO not only built a website to communicate the details of its myFICO project, but recently launched the “myFICO App” for consumers. Both are designed in a manner that allows consumers to easily get access to the features of FICO Score as well as how its scoring algorithm works. In the remainder of this

\(^2\) See myFICO booklet: “Understanding FICO Scores,” p. 3.
\(^3\) Source: the website of myFICO, https://ficoscore.com/education/.
\(^4\) See: https://www.myfico.com/credit-education?_ga=2.268952309.54042422.1555075673-650669123.1554110349
section, I aim to analyze FICO’s attempted algorithmic transparency from the perspective of informational disclosure.

I base the analysis mainly on the “Credit Education” section of the official website of myFICO, and a self-help booklet titled “Understanding FICO Scores,” downloaded from the site. Both provide detailed information about how the FICO Score works. To evaluate the extent of FICO’s algorithmic transparency, I follow Hurley and Adebayo’s proposal that three criteria should be met to for a credit scoring algorithm to be regarded as transparent. That is, the credit scoring entities must publicly publish information regarding.

1. The categories of data collected,
2. The sources and techniques used to acquire that data, and
3. The specific data points that a tool uses for scoring (Hurley & Adebayo, 2016, 204, 213).

First, the FICO Score makes clear which categories of data are used for the credit scoring calculation. As explained on the website, FICO Score considers four types of information in credit reports: personally identifiable information (PII), credit accounts, credit inquiries, and public record and collections. Personally identifiable information includes individuals’ “name, address, Social Security Number, date of birth and employment information.” Credit accounts consist of credit limit, account balance, payment history, and when individuals open their account (credit card, auto loan, etc.) Credit inquiries are the records of lenders’ checking individuals’ credit report within a period of time. Public record and collections contain the information about individuals’ bankruptcies and overdue debt gleaned from state and county courts.

Second, the FICO booklet illustrates how these sources of information are collected. As claimed, a FICO Score is solely calculated on the data aggregated from credit reports at the three major credit bureaus: Experian, TransUnion, and Equifax. Credit reports are the historical information recorded by credit bureaus to show how well a credit consumer has managed their credit. When applicants apply for credit, the lenders would report the updated applicants’ credit files (including credit data and other personal information) to credit bureaus. What FICO’s scoring algorithm does is to calculate the credit files into scores which are used to know applicant’s creditworthiness.

Third, the FICO website clearly explains the specific data points used for scoring which are listed by order of relative significance. According to the explanations, five data points account for a FICO Score: an individual’s payment history, amounts owed, length of credit history, new credit, and credit mix. Payment history is to measure how an applicant has paid past bills on time. Amounts owed is to gauge the amount of credit an applicant is using and how much debt is owed. Length of credit

---

5 Source: https://www.myfico.com/credit-education/whats-in-your-credit-score
6 See myFICO booklet: “Understanding FICO Scores”, p. 10.
7 See: https://www.myfico.com/credit-education/whats-in-your-credit-score
history is to consider how long an applicant’s credit accounts have been established. New credit is to analyze the characteristics of recently opened accounts. Credit mix takes into account an applicant’s different kinds of credit accounts being used or reported. More than that, the FICO website also explains how important for each category is in determining the calculation of consumers’ credit scores. The point of payment history accounts for 35% of a FICO Score. Amounts owed accounts for 30%. Length of credit history constitutes 15% of the score, while credit mix and new credit each contribute 10%.8

From an informational perspective, FICO Score’s disclosures meet all of basic requirements suggested by Hurley and Adebayo: the categories of data collected, the sources and techniques used to acquire that data, and the specific data points that a tool uses for scoring. That is probably why Hurley and Adebayo regard the FICO Score as a model of regulating black-boxed algorithms for other automated decision-making systems in the Big Data era (Hurley & Adebayo, 2016).

However, not everyone is satisfied with the extent of the disclosure offered by FICO, or by the way in which it is done. Danielle Citron and Frank Pasquale, for example, criticize FICO’s disclosing the inner workings of its algorithm as being far too general. The FICO Score has indeed explained the relative weight of five categories in its algorithm, but it does not “tell individuals how much any given factor mattered to a particular score” (Citron & Pasquale, 2014, 17). Some explanatory phrases like “balance owed” and “missed payment” are commonplace for those with both high and low credit scores. As for the category of “amounts owed”, for instance, no one knows whether using 25% of one’s credit limit is better or not than using 15% (Citron & Pasquale, 2014, 11). Therefore, FICO’s scoring algorithm is still highly opaque since “scored individuals unable to determine the exact consequences of their decisions” (Citron & Pasquale, 2014, 17).

Citron and Pasquale thereby propose that the FICO Score should provide more specific information about its algorithm in order that consumers are able to make the optimal credit utilization strategy. They suggest an “interactive modeling” to de-mystify algorithm: credit bureaus are advised to build a digital interface that “lets individuals enter various scenarios” to estimate how different decisions will affect their credit scores and provide “individuals more of a sense of how future decisions will affect their evaluation” (Citron & Pasquale, 2014, 29).

Despite their differences on how much the algorithm should be transparent, both Hurley and Adebayo and Citron and Pasquale are committed to an informational account of algorithmic transparency. On this account, it is the amount of information and its usefulness to potential users which determine whether the workings of a given algorithm are sufficiently transparent. Nonetheless, if FICO’s scoring system is a form of disciplinary surveillance, where consumers are ready (or self-disciplined) to improve their scores, the disclosed data points will just tell people (in ever greater detail) which norms to follow. For Citron and Pasquale, this disciplinary effect may be even stronger as their proposal of “interactive modeling” can make consumers more actively engaged in understanding and obeying the norms. In the

---

8 Ibid.
following two sections, I will take a closer look at the disciplinary power which is operated through the informational disclosure of algorithm on the part of FICO.

4 The Disciplinary Power of Credit Surveillance

Credit evaluation is an effective disciplinary surveillance, a tool to normalize citizens, much like the more well-known tools of laws and morality (Lauer, 2017). Traditional credit surveillance often operates its disciplinary power through punishments. The principle is to single out and exclude those known to be likely to default on their debts, the threat of which can discipline credit seekers to repay timely (Burton, 2008, 115). Contemporary credit surveillance operates in a more complicated way with the help of massive data collection. But the general idea is the same: to discipline people, or in Foucault’s term, to “train their soul” (Foucault, 1977).

Credit bureaus often collect so much data that they can be seen almost as all-seeing agencies that constantly surveil consumers’ behavior. In the past, credit information was scattered and creditors often relied on rumors and hearsay from credit applicants’ friends or neighbors to make decisions.9 It was only in the 1840s that recognizably modern credit reporting firms emerged (Lauer, 2017, 6). They gathered and standardized the hearsay regarding the applicants and then kept the relevant information in massive ledgers. Credit files at that time often contain “the intimate details of one’s domestic arrangements, personality, health, legal and criminal history, and job performance, and sometimes one’s physical appearance” (Lauer, 2017, 10). As Josh Lauer finds, the credit bureaus had mined so rich and reliable data that “local police and government often turned to the credit bureau for help” (Lauer, 2017, 10).

This process of extensive data collection itself can exert disciplinary effects over consumers since the mere thought that they may be monitored by credit bureaus would likely change their behavior. Besides such direct disciplinary impacts, the considerable collection of data is the source to exercise a stronger disciplinary power through sorting and punishments.

With massive data collection, each individual can be targeted to give an independent credit evaluation. Even if both positive and negative information are gathered, credit bureaus are more focused on collecting the negative ones (such as financial distress, gambling, or drinking) to judge people’s creditworthiness. All these negative items will be counted as risks in different weights according to normative financial expectations. For instance, if borrowers are alcoholics, or they repay bills slowly, they are more likely seen as a greater risk (Burton, 2012). Credit surveillance thereby sorts out “bad” credit consumers from “good” ones, which is similar

---

9 In 1748, Benjamin Franklin wrote a famous letter Advice to a Young Tradesman, in which Franklin vividly showed how creditor may actively surveil his debtors, observing and hearing for proof of honesty or vice: “The sound of your hammer at five in the morning, or eight at night, heard by the creditor, makes him easy six months longer; but if he sees you at the billiard-table, or hears your voice at a tavern, when you should be at work, he sends for his money the next day” (Franklin, 1748). Source: https://founders.archives.gov/documents/Franklin/01-03-02-0130
to what Foucault describes as a dividing practice of panopticon: “the ‘bad’ are separated from the ‘good’, the criminal from the law-abiding citizen, the mentally ill from the normal” (Burton, 2012, 114).

Consumers are aware of the effects of a poor credit rating, which encourages them to behave in ways that credit bureaus will approve of. Those who behave “badly” face exclusion. Blacklisting is a common way for such exclusion strategy. A blacklist is only a catalog of names used to record those who fail to make repayments. Credit bureaus use this list to figure out those individuals with the worst credit risks. Blacklisting is a negative system, the goal of which is not only to identify but also to quarantine and punish those delinquent persons and slow-payers. Sometimes, such lists are even published by the media to let the public know. In doing so, the individuals who struggle to pay would feel embarrassed and ashamed, which may force them to repay loans more efficiently. With this mechanism of blacklisting, consumers’ behavior can be more effectively disciplined, just like “the overseer’s book of penalties, which disciplined the worker” (Lauer, 2017, 11).

Hence, the capacity of sorting and its ensuing punishments can arouse a feeling of fear in credit customers, which will enforce them to obey the rules set up by credit bureaus. This disciplinary effect can be even amplified if a culture of credit has already established in society where creditworthiness is often seen as a moral character, which reflects one’s honesty, integrity, and responsibility. In this society, people with denied credit are expected to feel guilty because that credit denial means they are morally bad—being dishonest or irresponsible (Lazzarato, 2012, 30; Graeber, 2011, 121; Krippner, 2017, 10). This disciplinary power of credit surveillance has long been noted by credit bureaus. As a credit expert observed in the early twentieth century, “I know of no single thing, save the churches, which has so splendid a moral influence in the community as does a properly organized and effectively operated credit bureau” (Lauer, 2017, 127).

Admittedly, the disciplinary power of credit surveillance is not new, as many scholars have studied on the disciplinary effects of credit systems (Burton, 2008; Krippner, 2017; Lauer, 2017; Lazzarato, 2012). Nevertheless, some literature has argued that the Foucauldian disciplinary power of credit surveillance may have declined in digital society, since surveillance itself has experienced a post-panoptic turn (Langely, 2014; DuFault & Schouten, 2020). It is on this basis that Gilles Deleuze criticizes Foucault, arguing that in the computerized age the disciplinary society would be replaced by the society of control (Deleuze, 1992; Gane, 2012; Haggerty & Ericson, 2000).

That is perhaps why the credit scoring system has not received the critical attention it deserves. Instead of as a manifestation of power, it is seen only as an automated decision-making technology which does not seek “to direct or guide the individual’s decision-making processes” (Yeung, 2017, 121). This view distinguishes credit scoring algorithms from algorithmic recommendation systems, like those used by, for example, Amazon and Fitbit, which apply their algorithms to generate advice for their users and even manipulate them into buying certain products. Credit scoring algorithms are thus merely used by lenders “to differentiate, sort, target and price customers in terms of risk” rather than to discipline credit consumers to be responsible for making their repayments (Langely, 2014, 7).
This paper may be among the few pieces of literature that tries to make clear why the Foucauldian analysis is still crucial for understanding credit surveillance in digital society. In the remainder of this section, by studying FICO Score System, I provide two arguments to demonstrate that even in digital era the disciplinary power of credit surveillance is still relevant.

First, the FICO Score is still fundamentally a disciplinary system based on reward-and-punishment mechanism. In the era of credit scoring, creditors often set up a threshold to help lenders to decide whether to approve one’s credit application and in what terms and interest rates. The threshold is not fixed, varied according to different lenders, but FICO does provide a chart showing what a good FICO Score in a particular range means across the US population. This FICO’s scoring system can thereby still discipline a consumer by simply the threat of rejecting credit application or offering disadvantaged interest rates.

Besides the threat of a punishment, the FICO Score tends to become more and more an incentive-based system. As claimed, FICO’s algorithm does not follow the “knock out rules” that “turn down borrowers based solely on a past problem”; instead, it “weighs all of the credit-related information, both good and bad, in credit report.” In this new policy, one’s credit score can always be improved, so credit users are motivated to overhaul their credits persistently. This reward-driven feature is explicitly reflected throughout the FICO website, which is flooded with rosy words, pictures, and videos to show how a high FICO Score can help individuals reap plenty of perks.

Second, FICO’s scoring system is still basically about evaluating individuals’ character, in which financial and moral qualities cannot be untangled. It is right that computerized credit scoring has depersonalized credit evaluation and creditworthiness has been calculated as credit risks. But as Lauer shows, this scoring process “cannot eliminate the underlying morality of creditworthiness” (Lauer, 2017, 21). The judgment of one’s character may have been replaced by the language of credit risk, but the high or low risk still reflects people’s characters such as integrity or responsibility.

This is clearly revealed in the fact that credit scores are widely applied in some non-financial areas in the USA. For example, the Society of Human Resources Management (SHRM) conducted a survey showing that about 60% of employers in America perform credit background checks for their job candidates. Insurance

---

10 Source: https://www.myfico.com/credit-education/credit-scores/how-scoring-helps-you
11 For example, the headline writes, “Don’t let poor credit stand in the way of achieving your dreams.” There are also several videos on the website describing how people with high FICO Score can “save consumers thousands on a car loan or mortgage and give you access to the best credit cards, higher credit limits and more.” In the booklet of FICO’s algorithm, a specific story is narrated to emphasize the big difference between an individual with 620 FICO Score and a person with 760 FICO Score: it “can be tens of thousands of dollars over the life of a loan.” See: the booklet Understanding FICO Scores, published by FICO.
12 This situation may have been improved in recent years. As reported by the National Association of Professional Background Screeners and HR.com, about 25% of HR professionals still conduct credit checks on their job candidates. See: https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/pages/creditbackgroundchecks.aspx.
companies now use credit checks to decide which insurance rates should be assigned to drivers, houseowners, and even patients (Langenbucher, 2020). There is even a dating website, which promotes itself as a place “where good credit is sexy,” trying to match subscribers with credit scores. In these cases, credit scores are used not only to gauge the likelihood of financial default, but also to judge one’s “honesty, responsibility, and overall tendency toward conformity or friction” (Lauer, 2017, 22). As John Espenschied, the owner of Insurance Brokers Group, explains: “You can have a clean driving history with no driving violations or accidents, but still be penalized for poor credit, which means higher insurance premiums.” This happens because drivers with poor credit are often considered to be less responsible, which may pose more risks in driving cars.

Credit scores correlate with the general character of responsibility, which is also reflected on the website of FICO Score, where it often reminds people to identify higher credit score with responsible “healthy” behavior. For example, FICO repeatedly suggests that “to have higher FICO scores is to demonstrate healthy credit behaviors over time.” It has other similar claims: “Your FICO Scores are a vital part of your financial health” and “It’s never too late to reestablish healthy credit management habits.” Sometimes, the FICO Score even describes repairing bad credit as something like “losing weight.” In doing so, a higher FICO Score is framed as a desirable thing attained through the hard work of self-discipline. If people do not have good FICO Scores, the ensuing judgment will follow that people are irresponsible because they lack self-control and fail to keep themselves “healthy.”

To sum up, the FICO Score System intends to do more than simply calculating the risks of credit users; instead, it disciplines them according to some expected norms of being responsible credit consumers. What are those norms? And how does the process of normativity work? This is where the algorithmic transparency comes into play.

5 Normativity Through Algorithmic Transparency

The FICO Score System operates its disciplinary power to encourage consumers’ compliance with expected norms. Such a process of normativity needs algorithmic transparency to not only display but legitimate the norms in order to normalize consumers more effectively. In other words, algorithmic transparency itself has become a crucial part of disciplinary technique.

Normativity is at the heart of all disciplinary systems, as disciplinary power often exercises over people to make them conform to certain norms (Foucault, 1977, 177, 183). The norm, in general, is seen by Foucault as a new form of “law,” which defines and represses behavior that the judicial laws do not cover (ibid., 178). Institutions such as workshops, schools, and armies, as Foucault shows, have various

---

13 See the dating website: https://www.creditscoreddating.com/
14 See: https://www.insure.com/car-insurance/bad-credit-history.html
15 Source: https://ficoscore.com/education/
16 See myFICO booklet: “Understanding FICO Scores”, p.22.
17 Source: https://www.myfico.com/credit-education/improve-your-credit-score/
norms concerning the “time (lateness, absences, interruptions of tasks),” “activity (inattention, negligence, lack of zeal),” or “behavior (impoliteness, disobedience),” etc. (*ibid.*, 178). Here is a glimpse of some typical norms in an orphanage:

On entering, the companions will greet one another.
On leaving, they must lock up the materials and tools that they have been using and also make sure that their lamps are extinguished.
It is expressly forbidden to amuse companies by gestures or in any other way.
They must “comport themselves honestly and decently.”

… (*ibid.*, 178)

For Foucault, these norms are not arbitrary but are established according to some believed “truth,” which are often narrated by authoritative voices of scientific, medical, religious, or coercive institutions such as the police, schools, and prisons (Waitt, 2005, 174). The authoritative knowledge is given by various “experts” who define “truth” about “normality” and “deviance,” which are used to legitimate a particular set of norms. These norms become objectified when individuals regard them as natural and necessary, accepting them without critical analysis (Taylor, 2009, 52). When the objectification happens, people would be internally normalized, or self-disciplined, “into a certain frame of mind” that makes them conform to the norms even without outside enforcement (Forst, 2017; Hayward, 2018).

Admittedly, the FICO Score System is not a traditional institution in the same sense that schools, prisons, and armies are. The FICO Score seems more like what Deleuze calls “fluid” existence that there is no fixed shape but only numbers in and numbers out (Deleuze, 1992; Galič et al., 2017). But in fact, the FICO System does have its relatively fixed shape. The official website of myFICO has a particular “Credit Education” section, a place to educate people what FICO Scores are and how they calculate. Moreover, myFICO app is deployed in a way similar to personal fitness tracking devices like Fitbit, which helps people “be smart about (their) FICO Scores anywhere, anytime, and it’s fun to boot.”⁴⁸ Seen in this light, FICO Scores function as a type of “digital institution” that does have disciplinary power. In the remainder of this section, I will specifically show how FICO’s algorithmic transparency functions as not only a display of norms but an objectification of them.

### 5.1 Algorithmic Transparency as a Display of Norms

Transparency about how FICO Scores are calculated let consumers know not only how the algorithm works but what norms should be followed. In the past, the inner workings of a credit scoring system were often unknown to consumers and the scoring process operated like a black box (Pasquale, 2015). This opacity greatly limited the disciplinary effect of credit surveillance: credit users could not know what credit bureaus expect them to do to improve creditworthiness, so they would simply

---

⁴⁸ Source: [https://www.fico.com/en/newsroom/fico-introduces-free-app-for-iphone-and-ipod-touch-04-08-2010](https://www.fico.com/en/newsroom/fico-introduces-free-app-for-iphone-and-ipod-touch-04-08-2010)
passively wait to be evaluated (Creemers, 2018, 27). What algorithmic transparency does is to open the black box, letting people know what the rules are so that people can actively participate in the disciplinary system.

As shown, the FICO Score System discloses three types of explanations about how its algorithm works: “1) the categories of data collected, 2) the sources and techniques used to acquire that data, and 3) the specific data points that a tool uses for scoring” (Hurley & Adebayo, 2016, 204, 213). These disclosed explanations may seem to be scientific and objective information about FICO’s scoring algorithm. Nonetheless, the information is transformed into normative rules when arranged intentionally in a normative discourse. For example, with the informational disclosure of algorithm, the FICO Score clearly explains the specific data points used for scoring and all the data points are listed by order of relative significance. Based on these data points, FICO draws the characteristics of so-called The FICO High Achiever, referring to what it is like to be a consumer from the highest range of the scores:

**Payment History**: About 96% have no missed payments at all; Only about 1% have a collection listed on their credit report; Virtually none have a public record listed on their credit report.

**Amounts Owed**: Average revolving credit utilization ratio is less than 6%; Have an average of 3 accounts carrying a balance; Most owe less than $3,000 on revolving accounts (e.g., credit cards).

**Length of Credit History**: Most have an average age of accounts of 11 or more years; Age of oldest account is 25 years, on average.

The listed characteristics of “The FICO High Achiever,” derived from the different data points, have built an ideal model for credit consumers. This ideal model can subtly motivate credit applicants to identify themselves to the detailed characteristics of being a FICO High Achiever. This motivation works well because the FICO Score is essentially a disciplinary system, within which consumers tend to be ready to improve their credit scores under the reward-and-punishment mechanism. This normative power is explicitly found in those “tips” provided by the FICO System for people to upgrade their scores. Instead of just describing what the ideal model is, these “tips” as concrete norms are given to directly inform people how to identify themselves to the model. Some specific ones are quoted as follows:

Pay your bills on time.
Avoid having payments go to collections.
Keep balances low on credit cards and other ‘revolving credit.’
Don’t close unused credit cards in an attempt to raise your scores.

---

19 To be sure, the real situation can be more complex. For example, some rich people may not concern that much about their credit scores as they have plenty of cash. Some poor ones may not conform to what the norms show even though their credit scores are low. But generally, as a rational agent, within a disciplinary context, people tend to discipline themselves to obey the rules in order to avoid punishment and get more benefits.

20 Source: myFICO website [https://www.myfico.com/credit-education/improve-your-credit-score](https://www.myfico.com/credit-education/improve-your-credit-score)
Now, the seemingly scientific facts of FICO’s algorithm have successfully transitioned from *what it is* to *what should be*. By listing the ranking of the relative importance of the data points, for example, it has subtly established a set of standards about the behavior that are commendable and those that are not. These norms may seem to be less coercive—people can obey or dismiss them. But considering its strong system of reward and punishment, consumers as rational individuals will try to better their positions by making various systems to their advantage. They (especially for those who are financially vulnerable) would be more likely to be normalized in order to gain financial advantages. For instance, upon learning that “payment history” will be considered in FICO’s algorithm, individuals would tend to make prompt repayment to improve their credit scores. In doing so, the FICO Score System in fact inculcates various norms of responsibility into consumers’ mind.21

5.2 Algorithmic Transparency as an Objectification of Norms

More than displaying norms, algorithmic transparency also functions as a way to objectify these norms, making people treat these norms as natural rules. In doing so, people would be internally motivated or self-disciplined to abide by the norms with less critical analysis.

With informational disclosure of its algorithm, the FICO Score presents that its algorithmic scoring process is scientific and objective. FICO’s scoring algorithm is described as an impersonal calculation that only considers applicants’ financial data, not aggregating personal information such as name, education, age, and gender.22 Its algorithm is illustrated with a chart to show five components considered in a FICO Score, the relative weight of which is explained with exact number of percentages such as 35%, 30%, and 15%. In doing so, it tries to prove that its automatic calculation is based on scientific methods that apply to all credit consumers. Just as the FICO Score proudly announces on its website: “(FICO Scores) remove personal opinion and bias from the credit process,” providing “a scientific and objective evaluation.”23

That scientific narrative of how its algorithm works tries to convince people that the norms are objective: they are not derived from law or moral imperatives but a scientific knowledge of algorithm. These algorithm-based norms are rooted precisely in the explanations of FICO’s inner workings. For example, the norm “Pay your bills on time” is directly stemmed from the data point of “payment history.” “Keep balances low on credit cards” originates from the data point of “Amounts owed.” These algorithmic norms are far more easily objectified than moral or legal rules, since the former ones, packaged in the ideology of scientific algorithms, will be at once legitimised as objective. In doing so, people immediately regard these norms as natural facts, unwilling to challenge them.

21 FICO Score’s disciplining consumers’ behaviors by certain transparency has been proved to be effective in some empirical studies (Homonoff et al., 2021, 23–24).
22 See myFICO booklet: “Understanding FICO Scores,” pp.11, 24.
23 See myFICO booklet: “Understanding FICO Scores,” p.3.
Gradually, the objectification of algorithmic norms reframes people’s mind in a way of restraining their thinking of other possibilities. In fact, besides the objective narrative of algorithmic norms, there are at least three other alternative narratives: first, a narrative of arbitrariness. Citron and Pasquale point out that the aggregation and assessment of credit scores can be arbitrary, rather than commonly supposed to be objective. For example, the same person could get drastically different credit scores from different credit bureaus: “In a study of 500, 000 files, 29% of consumers had credit scores that differed by at least 50 points between the three credit bureaus” (Citron & Pasquale, 2014, 12). Evidence also shows that credit scoring systems can even “penalize cardholders for their responsible behaviors” (Citron & Pasquale, 2014, 12).

Secondly, there is a narrative of discrimination. FICO’s algorithm is announced to be impersonal, thanks to its not considering consumers’ personal information such as race, age, and gender. But technically, FICO’s Big Data analysis can indirectly know and calculate those personal characters based on large sets of financial data which is filled with social meaning. That is why credit scoring can often reflect social inequality which has already exists. For instance, the credit scoring process can have disparate impacts on women and minorities, the reason of which is even unknown for its algorithm designers (Citron & Pasquale, 2014, 14).

The third one is a narrative of unfairness. Some criticize the design of FICO’s scoring algorithm in the sense that it “unfairly benefit(s) lenders by allowing poor credit risks to saddle themselves with debt.” For example, payment history and debt owed weigh 65% in total of the FICO Score, while opening new credit only takes 10% of importance. This design which emphasizes the past credit records means that poor-risk users can hardly improve their FICO Scores due to their initial bad credit history. They have to be charged with higher interest rates. Higher interest rates, however, can lead to a greater accumulation of debts that can in turn lower down credit users’ FICO Scores. This is a vicious cycle in which a poor credit risk can easily be trapped, where they are allowed “to build debt but never be able to pay it all off.”

All these three alternative narratives can challenge the legitimacy of FICO’s algorithmic norms. If the scoring algorithm can be arbitrary, then it becomes useless for consumers to obey the norms. If the algorithm can be discriminative, the norms derived from the algorithm are not objective either. If consumers are disciplined into an unfair direction, normalization will turn out to be a way of manipulation.

However, when the objectification of norms happens, it may block people’s doing such critical thinking. When FICO’s norms are identified with objective rules, they may view these norms as merely scientific facts, instead of normative rules or standards of behavior. In doing so, the subjects become more easily to accept and obey the FICO’s norms. This uncritical self-discipline caused by

---

24 See: https://ficoforums.myfico.com/t5/Understanding-FICO-Scoring/The-History-of-Credit/m-p/199357_ga=2.64275692.256469586.1555773033-650669123.1554110349

25 Ibid.
objectification appeals to a cognitive mechanism rather than an institutional one. It is not derived from institutional mechanisms like FICO’s constant rewards and punishments, in which people are so ready to be normalized to improve their FICO Scores that they may not be inclined to suspect the established norms. Instead, the objectification of algorithmic norms will undermine individuals’ cognitive capacity for critical thinking, leading to a situation where people follow the norms only because of ideological conditioning.

The paper points out the hidden disciplinary power of credit scoring systems, but this does not mean that consumers are necessarily subject to that disciplinary power all the time. The goal of the paper is to unpack the sociotechnical discourse that tends to turn individuals into the objects of rating systems according to some prescribed norms. That means that individuals subject to an algorithmic credit rating are more likely to be subject to its disciplinary power. For example, in theory, American consumers can freely choose to or not to conform the norms defined by FICO’s scoring algorithm. But a favorable credit rating is crucial for Americans to pursue some important goals, like purchasing a home or car, to start a new business, and to seek higher education (Citron & Pasquale, 2014; Hurley & Adebayo, 2016). As a result, consumers as rational individuals would tend to normalize their behavior in ways to make themselves appear as responsible consumers in order to maximize their FICO scores, especially for those who are financially vulnerable.

For the three alternative narratives, I also acknowledge that they may not necessarily happen in actual operation of credit bureaus. For example, arbitrariness can be generally limited by the market, considering that a lender turning the FICO scores upside-down would have the risk of going bankrupt. Overt discrimination can also be restrained to some extent by the rise of purchasing power of some groups that have long been seen as financially disadvantaged, such as women and black people (The Economist, 2021). As for the unfairness, it can also be limited by how a credit market must work, where encouraging paying bills on time and punishing those who do not seem to be required for credit industry. However, admitting all these, what I am concerned about is that through algorithmic transparency, people may be disciplined in a way that they would be more likely to only focus on the scientific and objective narrative of its algorithm, ignoring other alternative narratives.

6 The Manipulative Potential of Algorithmic Transparency

Using the case of FICO’s scoring algorithm as an example, I have explained how algorithmic transparency can act as a disciplinary technique that conditions individuals’ behavior. This power account of algorithmic transparency reminds us of how disclosed information may nonetheless be used to manipulate our behavior. However, this does not mean that algorithmic transparency in general necessarily leads to manipulation. In this section, I will argue that algorithmic transparency may have manipulative potential when it is used in a context of asymmetrical power relations.
It is true that a general notion of “manipulation” is not always negative. Some may even argue that smiling or telling a joke in a conversation can be seen as a form of manipulation (Momin, 2021). However, as mentioned above, this paper follows a critical understanding of manipulation given by Susser et al., (2019, 3) which posits that manipulation works by “exploiting the manipulee’s cognitive (or affective) weaknesses and vulnerabilities in order to steer his or her decision-making process towards the manipulator’s ends.” Manipulation is morally objectionable because it exploits individuals’ vulnerabilities, and directs their behavior in ways that are likely to be to the benefit of the manipulator. This treats the manipulee as mere means, rather than an end in themselves. A typical example is that some companies attempt to detect and exploit users’ emotionally vulnerable state to sell more products and services (Petropoulos, 2022; Susser et al., 2019).

In the context of algorithmic transparency, I do not argue that the informational disclosure of algorithms necessarily manipulates one’s behavior, but rather that such disclosure is potentially manipulative when it is embedded in an unequal power structure. For example, in a commercial environment, generating a profit is the priority of corporations and consumers’ true interests can be disregarded by companies (Hardin, 2002). Algorithms are often designed and deployed by commercial entities whose main goal is to generate ever-larger profits. When users interact with algorithm-driven systems owned by companies, there is often the risk that companies can manipulate “users towards specific actions that are profitable for companies, even if they are not users’ first-best choices”26 (Petropoulos, 2022). In such a commercial environment, algorithmic transparency can become a deliberate strategy used by some companies to manipulate individuals’ behavior to make profits.

In the case of the FICO Score, such a manipulation through algorithmic transparency is subtle. Consumers may be unaware of its manipulative potential because it is interwoven with its potential to empower users to make “better” financial decisions. The FICO Score, as claimed, is to help individuals “manage their credit health.”27 People with “healthy” credit scores can indeed benefit themselves: they may be more likely to get loans or have more favorable interest rates. That is why some have suggested that FICO Score is a “personalized, quantified, dynamic measure of creditworthiness” that can be used by people to “track their progress over time, analogous to the role of a Fitbit in encouraging exercise” (Homonoff et al., 2021, 23–24). In other words, like Fitbit which is to keep people responsible for their body health, FICO Score can be used to foster responsibility to maintain an individual’s financial health.

Despite these benefits, however, FICO Score still has the possibility to manipulate individuals’ behavior. This paradox is similar to the manipulative practices of some health apps identified by Marijn Sax (2021). He argues that these for-profit health apps are often touted as “tools of empowerment” (345), and can be said to empower their users by making them more efficiently discipline themselves to enjoy

26 See the full text here: https://tinyurl.com/2p89uc52
27 Source: https://www.fico.com/en/newsroom/fico-introduces-free-app-for-iphone-and-ipod-touch-04-08-2010
a healthy life that is seen as fit by themselves. However, these apps can also potentially manipulate users’ behavior by “targeting and exploitation of people’s desire for health” (ibid., 346). Sax demonstrates that these apps are presented as tools to “optimize the health of the users, but in reality they aim to optimize user engagement and, in effect, conversion” (ibid., 345, emphasis in original). That means that those for-profit health apps are designed in a way that they exploit their users’ “natural desire for health” to make them “spend more time and possibly more money on health apps” than they may really want to (ibid., 350).

Notedly, the FICO Score, like Fitbit and other for-profit health apps, is not only used to guide people toward a “healthier” credit score, but also a commercial tool for lenders to make profits. As explicitly announced by itself, FICO’s scoring system is aimed at helping lenders “increase customer loyalty and profitability, reduce fraud losses, manage credit risk, meet regulatory and competitive demands, and rapidly build market share” (ibid., 346) (emphasis added). Hence, there is room for the credit scoring system to manipulate individuals’ behavior in a direction that may not be in the best interests of credit consumers. Banks, for example, try to balance “the interest rate they receive on the loan and the riskiness of the loan” in order to make the most of profits (Stiglitz & Weiss, 1988, 393). Ideally, banks would like to raise the interest rate as high as possible to maximize their revenue. But in practice, the rise in interest rate also means the increase in debtors’ average “riskiness” for default, which will in turn decrease the banks’ profits.

To reduce the average riskiness, FICO Score can help banks not only figure out who is more likely to repay, but also provide a disciplinary tool that can “directly control the actions of the borrower” (Stiglitz & Weiss, 1988, 393–394). As shown before, the FICO Score tries to discipline people to be “responsible” credit users through a deliberative strategy of algorithmic transparency. The responsible behaviors refer to some consuming characteristics, such as “pay bills on time,” “keeping balances low on credit cards,” and “avoid bankruptcy and collections,” which are exactly what lenders want. If credit users, especially those with poor credit risk, can follow those guidelines, lenders can effectively reduce the riskiness of rising interest rates. Banks, for instance, can loan money to some financially vulnerable individuals by imposing high interest rates with less worry of defaults, since those consumers who are disciplined to be “responsible” will tend to “pay bills on time” and may even feel guilty if they fail to repay the loan.

In this light, the FICO Score can potentially manipulate credit users’ behavior in a direction that ensures a lower risk of default, so that the lenders can set interest rates as high as possible to maximize their profits. Similar to some for-profit health apps’ exploiting users’ natural desire for health, such FICO Score targets and exploits consumers’ desire for financial or credit “health.” At first glance, it may appear that using FICO Score is beneficial to some credit users, but having those benefits does not mean that the FICO Score really cares about credit users’ true interests. As mentioned in the last section, the design of the FICO Score can potentially benefit lenders more than borrowers. For those with a high credit risk, their true interests are

---

28 Ibid.
not about being responsible or “pay bills on time,” but about how to get rid of the “trap” of debts triggered by the FICO Score System. If FICO Score really promotes the credit health of consumers, as it announces, its design should ensure that credit users and lenders benefit more evenly, rather than simply manipulating credit users to select inferior choices that may be more beneficial for lenders.

Admittedly, it is often hard to achieve algorithmic transparency without manipulation, since asymmetrical power relations are deeply rooted in our scored society. Just as Citron and Pasquale show, in their discussion of the scored society, “the realm of management and business more often features powerful entities who turn individuals into ranked and rated objects” (Citron & Pasquale, 2014, 3). Under such an unequal power structure, there is often room for the powerful entities to potentially manipulate individuals’ behavior through algorithmic transparency. Despite the difficulties, we still should not give up challenging the unbalanced power structure to not let algorithmic transparency become a manipulative strategy. Like governing other algorithmic systems, a set of policies are required to ensure that algorithmic systems should not use manipulative strategies to undermine consumers’ true interests. Such regulations should make sure not only that their algorithms are transparent, but also that their deliberative disclosure of their algorithm itself is not manipulative. Additionally, we need to encourage more public discussion and education about the manipulative risks of algorithmic systems, so that consumers are more reflective and critical about the systems which shape the choices they make.

What’s more, a possible way of reforming the disciplinary power exercised by algorithmic scoring systems could be to design them in ways that induce less stress and more tolerance in those that are subject to them. For scoring systems, it is often the fear and anxiety of punishments that prompt subjects to discipline themselves to maximize their ratings. The more severe the punishments, the more likely people are subject to the pressure of disciplinary power. Hence, to reduce its disciplinary effects, we could consider incorporating some notion of forgiveness into our algorithmic systems, tolerating more mistakes, inconsistencies, and unpredictability that are an inherent part of the human condition. For example, the Chinese Social Credit System is designed in a relatively unforgiving manner. As Chinese authorities bluntly stress, the Social Credit System is to make sure “the trustworthy benefit at every turn and the untrustworthy can’t move an inch.” Under this principle, the so-called untrustworthy citizens are immediately sanctioned and face long-lasting and significant punishments (Ding & Zhong, 2021; Drinhausen & Brussee, 2021). Such an overly disciplinary system should include more tolerant elements, such as making its punishments less harsh or allowing citizens to freely opt out of the system, etc.

7 Conclusion

The idea of algorithmic transparency has become an important tool for pushing back against automated decision-making. The EU’s groundbreaking General Data Protection Regulation (GDPR) even includes “a right to explanation” regarding the inner working of the algorithmic decision-making process (Kaminski, 2019; Rochel, 2021). There is also a growing body of empirical research on this issue suggesting
that companies do not give detailed explanations or only sketchily describe their algorithms upon request (Dexe et al., 2022; Krebs et al., 2019; Sørøm & Presthus, 2021). The GDPR’s “right to explanation” and most of the existing empirical studies on this topic mainly focus on an informational version of algorithmic transparency—about how much information should be revealed about the inner workings of an algorithm.29 However, for a Foucauldian power analysis as shown in this paper, it may not concentrate on how much information has been disclosed, but is more concerned with whether there is an existing power imbalance. It may hold that such sketchily described algorithms can potentially be tools of manipulation in our scored society. If an automated system is inherently a disciplinary one, making its algorithm transparent may solely serve to increase the normalizing power behind it.

**Acknowledgements** I would like to extend my sincere thanks to Gerrit Schaafsma, Beate Roessler, Jeroen de Kloet, Marjolein Lanzing, and Marijn Sax for helpful comments on earlier drafts of this article. I also benefited from feedback from participants at Privacy Law Scholar’s Conference (PLSC) and The Society for Philosophy and Technology (SPT) conference. I gratefully acknowledge support from the China Scholarship Council (CSC) and the Amsterdam School for Cultural Analysis (ASCA) at University of Amsterdam.

**Author Contribution** The author confirms sole responsibility for the manuscript.

**Funding** The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the China Scholarship Council (CSC Grant: #201606360144).

**Data Availability** Not applicable.

**Declarations**

**Ethics Approval and Consent to Participate** Not applicable.

**Consent for Publication** Not applicable.

**Competing Interests** The author declares no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

---

29 My paper only mentions the GDPR and its “right to explanation” in a brief manner. Since the GDPR is highly relevant and rather complicated, perhaps for future study I could apply the power account of algorithmic transparency to make a particular analysis on the GDPR.
References

Ahmed, S. (2018). Credit cities and the limits of the social credit system. In AI, China, Russia, and the Global Order. Wright, N.D. (editor) http://nsiteam.com/social/wp-content/uploads/2019/01/AI-China-Russia-Global-WP_FINAL_forcopying_Edited-EDITED.pdf#page=63

Albu, O. B., & Flyverbom, M. (2016). Organizational transparency: Conceptualizations, conditions, and consequences. Business and Society, 58(2), 268–297.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media and Society, 20(3), 973–989.

Baum, K., Mantel, S., Schmidt, E., et al. (2022). From responsibility to reason-giving explainable artificial intelligence. Philosophy and Technology, 35(1), 1–30. https://doi.org/10.1007/s13347-022-00510-w

Beer, D. (2017). The social power of algorithms. Information Communication and Society, 20(1), 1–13.

Binns, R. (2018). Algorithmic accountability and public reason. Philosophy and Technology, 31(4), 543–556. https://doi.org/10.1007/s13347-017-0263-5.

Black, J. S., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it? Business Horizons, 63(2), 215–226.

Broussard, M. (2020). When algorithms give real students imaginary grades. The New York Times. https://www.nytimes.com/2020/09/08/opinion/international-baccalaureate-algorithm-grades.html

Buchanan, T. (2018). If... then: Algorithmic power and politics. Oxford University Press.

Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. Big Data and Society, 3(1), 1–12.

Burton, D. (2008). Credit and consumer society. Routledge.

Burton, D. (2012). Credit scoring, risk, and consumer lendingscapes in emerging markets. Environment and Planning, 44(1), 111–124. https://doi.org/10.1068/2Fa44150

Calo, R. & Citron, D. K. (2021). The automated administrative state: A crisis of legitimacy. Emory LJ, 70(4), 797–845. https://scholarlycommons.law.emory.edu/elj/vol70/iss4/1/

Chan, N. K. (2019). The rating game: The discipline of Uber’s user-generated ratings. Surveillance and Society, 17(1/2), 183–190.

Citron, D. K., & Pasquale, F. A. (2014). The scored society: Due process for automated predictions. Washington Law Review, 89(1), 1–34. https://digitalcommons.law.uw.edu/wlr/vol89/iss1/2

Clifford, R., & Shoag, D. (2016). “No more credit score”: Employer credit check bans and signal substitution. FRB of Boston Working Paper N. 16–10 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2836374

Creemer, R. (2018) China’s Social Credit System: An evolving practice of control. Available at SSRN: https://ssrn.com/abstract=3175792 or https://doi.org/10.2139/ssrn.3175792

Dai, X. (2018). Toward a reputation state: The Social Credit System Project of China. Available at SSRN: https://ssrn.com/abstract=3193577 or https://doi.org/10.2139/ssrn.3193577

Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. Philosophy and Technology, 29(3), 245–268.

De Laat, P. B. (2018). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? Philosophy and Technology, 31(4), 525–541.

De Laat, P. B. (2019). The disciplinary power of predictive algorithms: A Foucauldian perspective. Ethics and Information Technology, 21(4), 319–329. https://doi.org/10.1007/s10676-019-09509-y

Deleuze, G. (1992). Postscript on the societies of control. October 59, 3–7. http://www.jstor.org/stable/778828

Dexe, J., Franke, U., Söderlund, K., et al. (2022). Explaining automated decision-making: A multinational study of the GDPR right to meaningful information. The Geneva Papers on Risk and Insurance - Issues and Practice, 47, 669–697. https://doi.org/10.1057/s41288-022-00271-9

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. Digital Journalism, 3(3), 398–415.

Diakopoulos, N. (2020). Transparency. In M. Dubber, F. Pasquale, & S. Das (Eds.), Oxford handbook of ethics and AI (pp. 197–214). Oxford University Press.

Diakopoulos, N., & Koliska, M. (2016). Algorithmic transparency in the news media. Digital Journalism, 5(7), 809–828.

Diakopoulos, N. (2018). The algorithms beat, http://www.nickdiakopoulos.com/wp-content/uploads/2018/04/Diakopoulos-The-Algorithms-Beat-DDJ-Handbook-Preprint.pdf
Ding, X., & Zhong, D. Y. (2021). Rethinking China’s Social Credit System: A long road to establishing trust in Chinese society. *Journal of Contemporary China, 30*(130), 630–644.

Dobbie, W., Goldsmith-Pinkham, P., Mahoney, N., & Song, J. (2016). Bad credit, no problem? Credit and labor market consequences of bad credit reports. *The Journal of Finance, 75*(5), 2377–2419. https://doi.org/10.1111/jofi.12954

Drinhausen, K., & Brussee, V. (2021). China’s Social Credit System in 2021: From fragmentation towards integration, *MERICS China Monitor, 12*. https://merics.org/en/report/chinas-social-credit-system-2021-fragmentation-towards-integration

Drucker, S. J., & Gumpert, G. (2007). Through the looking glass: Illusions of transparency and the cult of information. *Journal of Management Development, 26*, 493–498.

DuFault, B. L., & Schouten, J. W. (2020). Self-quantification and the datapreneurial consumer identity. *Consumption Markets and Culture, 23*(3), 290–316.

Erasmus, A., Brunet, T. D. P., & Fisher, E. (2021). What is interpretability? *Philosophy and Technology, 34*, 833–862. https://doi.org/10.1007/s13347-020-00435-2

Estop, J. D. (2014). WikiLeaks: From Abbé Barruel to Jeremy Bentham and beyond (A short introduction to the new theories of conspiracy and transparency). *Cultural Studies Critical Methodologies, 14*(1), 40–49.

Etzioni, A. (2010). Is transparency the best disinfectant? *Journal of Political Philosophy, 18*, 389–404.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. Martin’s Press.

Flyverbom, M., Christensen, L. T., & Hansen, H. K. (2015). The transparency–power nexus: Observational and regularizing control. *Management Communication Quarterly, 29*(3), 385–410.

Forst, R. (2017). *Normativity and power: Analyzing social orders of justification, trans*. Oxford University Press.

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. (A. Sheridan, Trans.). Vintage Books. (Original work published 1975)

Franke, U. (2022). First-and second-level bias in automated decision-making. *Philosophy and Technology, 35*(2), 1–20. https://doi.org/10.1007/s13347-022-00500-y

Galič, M., Timan, T., & Koops, B. J. (2017). Bentham, Deleuze and beyond: An overview of surveillance theories from the panopticon to participation. *Philosophy and Technology, 30*(1), 9–37.

Gane, N. (2012). The governmentalities of neoliberalism: Panopticism, post-panopticism and beyond. *The Sociological Review, 60*(4), 611–634.

Graeber, D. (2011). *Debt: The first 5,000 years*. Melville House.

Haggerty, K. D., & Ericson, R. V. (2000). The surveillant assemblage. *The British Journal of Sociology, 51*(4), 605–622.

Hansen, H. K., & Flyverbom, M. (2014). The politics of transparency and the calibration of knowledge in the digital age. *Organization, 22*(6), 872–889.

Hardin, R. (2002). *Trust and trustworthiness*. Russell Sage Foundation.

Hayward, C. R. (2018). On structural power. *Journal of Political Power, 11*(1), 56–67.

Heald, D. (2006). Transparency as an instrumental value. In C. Hood & D. Heald (Eds.), *Transparency: The key to better governance?* (pp. 59–73). Oxford University Press.

Homonoff, T., O’Brien, R., & Sussman, A. B. (2021). Does knowing your fico score change financial behavior? Evidence from a field experiment with student loan borrowers. *Review of Economics and Statistics, 103*(2), 236–250.

Hurley, M., & Adebayo, J. (2016). Credit scoring in the era of big data. *Yale JL and Technology, 18*, 148.

Jauernig, J., Uhl, M., & Walkowitz, G. (2022). People prefer moral discretion to algorithms: Algorithm aversion beyond intransparency. *Philosophy and Technology, 35*, 2. https://doi.org/10.1007/s13347-021-00495-y

Kaminski, M. (2019). The right to explanation, explained. *Berkeley Technology Law Journal, 34*(1), 189–218. https://doi.org/10.15779/Z38TD9N83H

Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication and Society, 22*(14), 2081–2096. https://doi.org/10.1080/1369118X.2018.1477967

Kim, P. T. (2020). Manipulating opportunity. *Virginia Law Review, 106*(4), 867–935.

Kim, K., & Moon, S. I. (2021). When algorithmic transparency failed: Controversies over algorithm-driven content curation in the South Korean digital environment. *American Behavioral Scientist, 65*(6), 847–862.
Kossow, N., Windwehr, S., & Jenkins, M. (2021). Algorithmic transparency and accountability. Transparency International Anti-Corruption Helpdesk Answer. https://knowledgehub.transparency.org/assets/uploads/kproducts/Algorithmic-Transparency_2021.pdf

Krebs, L. M., Alvarado Rodriguez, O. L., Dewitte, P., Ausloos, J., Geerts, D., Naudts, L., & Verbert, K. (2019, May). Tell me what you know: GDPR implications on designing transparency and accountability for news recommender systems. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–6.

Krippner, G. R. (2017). Democracy of credit: Ownership and the politics of credit access in late twentieth-century America. American Journal of Sociology, 123(1), 1–47. https://doi.org/10.1086/692274

Langely, P. (2014). Equipping entrepreneurs: Consuming credit and credit scores. Consumption Markets and Culture, 17(5), 448–467.

Langenbucher, K. (2020). Responsible AI-based credit scoring—a legal framework. European Business Law Review, 31(4), 527–572.

Lauer, J. (2017). Creditworthy: A history of consumer surveillance and financial identity in America. Columbia University Press.

Lazzarato, M. (2012). The making of the indebted man: An essay on the neoliberal condition (J.D. Jordan, Trans.). Semiotext.

Lyon, D. (2007). National ID cards: Crime-control, citizenship and social sorting. Policing: A Journal of Policy and Practice, 1(1), 111–118.

Matzner, T. (2017). Opening black boxes is not enough: Data-based surveillance in discipline and punish and today. Foucault Studies, 23(2017), 27–45. https://doi.org/10.22439/fs.v0i0.5340

McKay, C. (2020). Predicting risk in criminal procedure: Actuarial tools, algorithms, AI and judicial decision-making. Current Issues in Criminal Justice, 32(1), 22–39.

Meijer, A. (2013). Understanding the complex dynamics of transparency. Public Administration Review, 73(3), 429–439. https://doi.org/10.1111/puar.12032

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. In Proceedings of the Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, New York, NY, USA, 279–288. https://doi.org/10.1145/3287560.3287574

Momin, K. (2021). Romantic manipulation: 15 things disguised as love. Bonobology. https://www.bonobology.com/romantic-manipulation/

Muldoon, J., & Raekstad, P. (2022). Algorithmic domination in the gig economy. European Journal of Political Theory, 14748851221082078. https://doi.org/10.1177/2F14748851221082078

myFICO Handbook. Retrieved 4 May 2020. https://www.myfico.com/credit-education-static/doc/education/myFICO_UYFS_Booklet.pdf

O’Connor, P. (2022). Coercive visibility: Discipline in the digital public arena. In P. O’Connor & M.I. Benta (Eds.), The technologisation of the social: A political anthropology of the digital machine (pp. 153–171). Routledge. https://doi.org/10.4324/9781003052678

O’Neil, C. (2016) Weapons of math destruction: How Big Data increases inequality and threatens democracy. Broadway Books.

Packin, N. G., & Lev-Aretz, Y. (2016). On social credit and the right to be unnetworked. Columbia Business Law Review, 2016(2), 339–425. https://doi.org/10.7916/cblr.v2016i2.1739

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Petropoulos, G. (2022, February 2). The dark side of artificial intelligence: manipulation of human behaviour. Bruegel-Blogs. https://www.bruegel.org/blog-post/dark-side-artificial-intelligence-manipulation-human-behaviour

Powell, A. B. (2021). Explanations as governance? Investigating practices of explanation in algorithmic system design. European Journal of Communication, 36(4), 362–375.

Prat, A. (2005). The wrong kind of transparency. American Economic Review, 95(3), 862–877.

Rochel, J. (2021). Ethics in the GDPR: A blueprint for applied legal theory. International Data Privacy Law.

Sax, M. (2021). Optimization of what? For-profit health apps as manipulative digital environments. Ethics and Information Technology, 2021(23), 1–17. https://doi.org/10.1007/s10676-020-09576-6

Schwerzmann, K. (2021). Abolish! Against the use of risk assessment algorithms at sentencing in the US criminal justice system. Philosophy and Technology, 34(2021), 1883–1904. https://doi.org/10.1007/s13347-021-00491-2

Springer
Smith, A. (2018). Franken-algorithms: The deadly consequences of unpredictable code. The Guardian. https://www.theguardian.com/technology/2018/aug/29/coding-algorithms-frankenalgos-program-danger

Sørum, H., & Presthus, W. (2021). Dude, where’s my data? The GDPR in practice, from a consumer’s point of view. Information Technology and People, 34(3), 912–929. https://doi.org/10.1108/ITP-08-2019-0433

Springer, A., & Whittaker, S. (2019). Progressive disclosure: empirically motivated approaches to designing effective transparency. In Proceedings of the 24th International Conference on Intelligent User Interfaces (pp. 107–120).

Stiglitz, J. & Weiss, A. (1988). Credit rationing in markets with imperfect information. American Economic Review, 71(3), 393–410. https://www.jstor.org/stable/1802787

Susser, D., Rössler, B., & Nissenbaum, H. (2019). Online manipulation: Hidden influences in a digital world. Georgetown Law Technology Review, 4(1), 1–45. https://doi.org/10.2139/ssrn.3306006

Susser, D. (2019). Invisible influence: Artificial intelligence and the ethics of adaptive choice architectures. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 403–408).

Taylor, D. (2009). Normativity and Normalization. Foucault Studies, 2009(7), 45–63.

The Economist (2021). Design bias is harmful, and in some cases may be lethal https://www.economist.com/leaders/2021/04/10/design-bias-is-harmful-and-in-some-cases-may-be-lethal

Tuffley, D. (2021). Love in the time of algorithms: Would you let your artificial intelligence choose your partner? The Conversation. https://theconversation.com/love-in-the-time-of-algorithms-would-you-let-artificial-intelligence-choose-your-partner-152817

von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. Philosophy and Technology, 34(2021), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0

Waitt, G. R. (2005). Doing discourse analysis. In I. Hay (Ed.), Qualitative research methods in human geography (pp. 163–191). Oxford University Press.

Weller, A. (2017). “Challenges for transparency,” in: ICML Workshop on Human Interpretability. https://doi.org/10.48550/arXiv.1708.01870

Yeung, K. (2017). ‘Hypernudge’: Big Data as a mode of regulation by design. Information Communication and Society, 20(1), 118–136.

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

Zavrhnik, A. (2021). Algorithmic justice: Algorithms and big data in criminal justice settings. European Journal of Criminology, 18(5), 623–642.

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in algorithmic and human decision-making: Is there a double standard? Philosophy and Technology, 32(2019), 661–683. https://doi.org/10.1007/s13347-018-0330-6

Zhang, C. (2020). Governing (through) trustworthiness: Technologies of power and subjectification in China’s Social Credit System. Critical Asian Studies, 52(4), 565–588.

Zou, S. (2021). Disenchanting trust: Instrumental reason, algorithmic governance, and China’s emerging Social Credit System. Media and Communication, 9(2), 140–149.

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

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