Factors Influencing Perceived Legitimacy of Social Scoring Systems: Subjective Privacy Harms and the Moderating Role of Transparency

Completed Research Paper

Carmen Loefflad **Mo Chen** **Jens Grossklags**
carmen.loefflad@tum.de mo.chen@tum.de jens.grossklags@in.tum.de
Technical University of Munich
Boltzmannstr. 3, 85748 Garching

Abstract

Technological advancements in recent years have enabled the spread of automated decision-making (ADM) systems. Social scoring systems are a specific instance of ADM system, using behavioral scores to encourage pro-social behaviors. Building on a survey following an experimental study, we present two structural equation models to determine the impacts of different levels of transparency on the perceived legitimacy of scoring systems, as well as on people's intention to comply with the system. The models are built on well-established theories highlighting procedural justice and outcome favorabilities as key determinating factors. Our results suggest that the determinants of perceived legitimacy are strongly shaped by the level of transparency. However, transparency elevates subjective privacy harms. Our findings add to the ongoing debate on the transparency of ADM systems, by identifying a trade-off between the elimination of outcome favorabilities in determining perceptions of legitimacy, and increased subjective privacy harms, weakening people's intention to comply.

Keywords: *human factors, experiment, social scoring systems, procedural justice, legitimacy, subjective privacy harms, compliance with social scoring systems*

Introduction

The developments in artificial intelligence have advanced the spread of automated decision-making (ADM) systems. These systems are mainly used for decision support, e.g., in hiring contexts (Li et al., 2021), or for assessing recidivism (Azeroual et al., 2020) and credit risks (Njuguna & Sowon, 2021). Social scoring systems are a specific kind of ADM system, which use behavioral data to calculate automated scores. Typically, social scoring systems try to achieve some form of behavioral change, as higher scores are associated with benefits to encourage certain types of behavior. Scholars refer to this practice as algorithmic regulation (O'Reilly, 2013; Yeung, 2018). For example, the large-scale Chinese Social Credit System purportedly aims to achieve a more trustworthy society (State Council, 2014); smaller scoring systems in Europe try to educate people on pro-environmental behaviors (Wolfangel, 2022). Hereby, the achievement of postulated goals is contingent on people's voluntary use of the system, and their voluntary compliance with the propagated behaviors (Yeung, 2018). However, people's decision to comply with the regulatory guidelines issued by a system is generally determined by the viewpoints they hold; to these counts first, the extent to which people perceive the deployed procedures of a system as just (*procedural justice*), and second, the extent to which people perceive the outcome they receive under the system, i.e., their score, as favorable (*outcome favorability*). Thirdly, the extent to which people consider a system legitimate is key for enabling people to adopt certain behaviors (*legitimacy*) (Brockner, 2002; Thibaut & Walker, 1975; Tyler & Jackson, 2013).

Existent work has focused on the determinants of people's trust in and the legitimacy of ADM systems (Araujo et al., 2020; Jacovi et al., 2021; Vereschak et al., 2021), as well as on people's perceptions of fairness of these systems (Lee et al., 2019; Wang et al., 2020). However, research investigating people's attitudes toward social scoring systems is scarce. With our work, we aim to explore the perceived legitimacy of scoring systems and people's intention to comply with them, given different levels of transparency regarding the scoring mechanism. The present paper focuses on a survey that followed an experimental study. The experiment investigated people's fairness-related behaviors in small-scale communities of strangers, to which an institutional scoring system was introduced, aiming at educating people to fair behaviors. The scoring treatments differed in the level of transparency given about the scoring mechanism. We present two structural equation models (SEM) to identify the determinants of the perceived legitimacy of social scoring systems, as well as people's intention to comply with the systems. The models are inspired by procedural justice theory (Thibaut & Walker, 1975), and the literature on the legitimacy of governance institutions (Mazerolle et al., 2013; Tyler & Jackson, 2013), as well as of ADM systems (K. de Fine Licht & de Fine Licht, 2020; Waldman & Martin, 2022). The models rely on three core perceptional variables: legitimacy, procedural justice, and outcome favorability. We further explore how the subjective privacy harms that are caused by the practices of digital scoring systems, which rely on a large-scale behavioral data collection, influence people's perceptions of social scoring systems, as well as their intention to comply. While the impacts of providing transparency on people's perceptions and intention to comply are tested under a confirmatory approach, our research specifically addresses two questions. *First, how do subjective privacy harms influence people's perceptions of social scoring systems, as well as people's willingness to comply with the system? Second, how does transparency alter the impact of subjective privacy harms on people's perceptions of social scoring systems, as well as on people's intention to comply?* Therefore, the scope of the present study is to investigate a system-level property, namely the impact of transparency on people's perceptions and compliance intentions. In addition, we use an experimental approach to study people's experiences with a social scoring system trying to incentivize one specific kind of behavior, namely fairness.

Our results suggest that procedurally just mechanisms are key for determining the perceived legitimacy of social scoring systems, as well as people's intention to comply, irrespective of outcome favorabilities. Procedurally unjust mechanisms, in contrast, increase the importance of outcome favorability in determining perceived legitimacy and the intention to comply, which is also referred to as *outcome favorability bias* (Wang et al., 2020). Importantly, whether scoring mechanisms are perceived as just is determined by the level of transparency. While related work proposes to incorporate a measure of "outcome favorability bias" into models assessing perceptions of ADM systems (Wang et al., 2020), our work suggests that the bias may be eliminated by increased levels of transparency. Yet, subjective privacy harms make a difference. They decrease people's satisfaction with their experience under a scoring system, and weaken people's intention to behave in line with the propagated behaviors. Surprisingly, these negative effects are only observed under a transparent mechanism. These results add another viewpoint to the ongoing debate about the "right" level of transparency in modern ADM systems (Ananny & Crawford, 2018; Burrell, 2016), by suggesting an interdependency between privacy and transparency in determining perceptions.

The remainder of the article is structured as follows: we first provide background about social scoring systems and procedural justice theory. Subsequently, we establish a formal research model of the determinants of legitimacy of social scoring systems, as well as people's intention to comply. The method section describes the scoring experiment and introduces the method of structural equation modeling. Subsequently, we present the results. Lastly, we discuss the implications of our results and conclude.

Background

Social Scoring Systems for Behavioral Regulation

Social scoring systems are applied across the globe and with different scales. A central characteristic of scoring systems is the use of behavioral scores to incentivize certain types of behaviors (Yeung, 2018). To these count pro-environmental (Wolfangel, 2022), or pro-social behaviors, such as showing increased trustworthiness (State Council, 2014). Many scholars attribute a great potential to social scoring systems in addressing society-wide problems (Cristianini & Scantamburlo, 2020; O'Reilly, 2013). Related research has shown that people are generally "not blind to the potential benefits of ADM" (Araujo et al., 2020). However, people tend to voluntarily comply with desired behaviors only if they view a system as legitimate (Lind & Earley, 1991; Lind & Tyler, 1988). Yet, social scoring systems inherently raise ethical considerations, casting doubt about their legitimacy. These systems are commonly described as "black boxes", which obscure decision-making principles. Further concerns are given by their potential disparate impact on disadvantaged groups of people, and their privacy-invading character (Citron & Pasquale, 2014; Zarsky, 2015). The critiques that have been raised on a conceptual level imply that once social scoring systems are characterized by opacity, invade people's privacy, or lead to a disadvantageous treatment of certain groups of people, people's views regarding the legitimacy of these systems may be negatively affected. Additionally, people's willingness to voluntarily use the system and comply with the desired behaviors may decrease.

Behavioral research regarding the relationship between specific properties of social scoring systems and people's perceptions of the system is scarce. However, since scoring systems are social systems, research regarding the factors that determine their use and perceived legitimacy is necessary to deepen our understanding of their potential benefits and harms (Ehsan et al., 2021). A widely discussed factor impacting people's perceptions of ADM systems is the level of transparency provided about their functioning (Ananny & Crawford, 2018; Burrell, 2016; Ehsan et al., 2021). One specific form of transparency is process transparency, which refers to the transparency level regarding how an evaluation model works (Drozdal et al., 2020). For example, giving people explanations about an algorithm constitutes a kind of process transparency. Understanding how a system works increases the interpretability of the system's mechanism, and helps people judge whether the system behaves in a way that is consistent with its purported motives (Rader et al., 2018). Transparency is closely related to perceptions of privacy, as transparency may lead people to develop an increased consciousness about the value of their data (Büchi et al., 2023). As such, transparency possibly contributes to fostering individuals' recognition of the level of privacy invasion that is associated with the data collection practices of ADM systems. In the context of our research, these observations suggest that higher privacy awareness, induced by increased transparency concerning the working principles of social scoring systems, may impact people's viewpoints of the system.

Procedural Justice Theory

Procedural justice theory (Thibaut & Walker, 1975) is concerned with people's assessments of and behavioral responses to governance institutions and authorities, such as courts or the police. Often, these authorities seek to ensure compliant or cooperative behaviors (Tyler & Jackson, 2013). Being subject to the decision-making of an authority, people typically receive a decision, which is referred to as their outcome. A key facet of procedural justice theory is to separate people's perception of the process leading to a certain outcome from people's subjective evaluation of the outcome, i.e., the so-called outcome favorability (Brockner & Wiesenfeld, 1996). A second key focus of the theory is to determine how reactions to a decision-making organ are shaped by perceptions of procedural justice. To these reactions count perceptions of legitimacy (Suchman, 1995), and people's intention to comply with the authority (Lind & Tyler, 1988). Procedural justice theory provides a suitable basis for analyzing people's perceptions of social scoring systems, because governance institutions and social scoring systems have similar objectives. First, both social scoring systems and governance institutions seek to incentivize certain types of behaviors. Second, they both render judgments about individuals based on their behaviors. Building on procedural justice theory, we aim to investigate the impact of different levels of process transparency on people's perceptions of social scoring systems. To these perceptions count people's perceived favorability of their experience with the system, perceptions of legitimacy, procedural justice, and subjective privacy harms. In addition, we examine the relationship between people's perceptions and their intention to comply with the system. In the following, we describe the perceptional variables and establish a formal research model.

Related Work and Research Hypotheses

For the purpose of ensuring a functioning society, governance institutions often pursue the establishment of a set of rules and laws (Hardin, 1993; Letki, 2006). People's compliance with these rules is key to accomplishing this goal. However, compliance with an institution is contingent on its legitimate status: only those who perceive an institution as legitimate tend to comply (Lind & Earley, 1991; Tyler, 2006b). Perceptions of legitimacy are determined by another set of factors, including the assessment of whether the deployed procedures are just and people's subjective favorability of a certain encounter with the institution (Mazerolle et al., 2013; Tyler & Jackson, 2013). In this section, we apply the rationales relating to the perceived legitimacy of institutional governance systems to the context of ADM systems. However, in today's digital age, people's willingness to adopt certain behaviors postulated by ADM systems may be sensitive to factors that arise uniquely due to the automated and data-driven character of decision-making processes. As ADM systems rely on a large-scale collection of behavioral data, people's subjective privacy harms may play a decisive role in determining perceptions or compliance intentions (Araujo et al., 2020; Thurman et al., 2019). In the following, we outline these factors and develop our hypotheses accordingly.

Perceived Procedural Justice and Perceived Legitimacy

Procedural justice refers to the perceived fairness of decision-making processes that lead to a certain outcome (Thibaut & Walker, 1975). Commonly, different elements determine the broad concept of procedural justice, including citizen participation, fairness and neutrality, dignity and respect, and benevolent motives of an authority (Alessandro et al., 2021; Brockner, 2002; Mazerolle et al., 2013; Tyler & Fagan, 2006, 2006).

The legitimacy of systems generally refers to the viewpoint that the actions of a system that operates in a society, which is characterized by specific values and norms, are appropriate (Suchman, 1995). However, legitimacy is a multifaceted construct; a central characteristic of popular legitimacy is the elicitation of a subjective need to contribute to the achievement of the desired goals of a system (Sunshine & Tyler, 2003). Further, legitimacy refers to the general view that the actions and decisions of a system are normatively aligned with people's own moral codes. Lastly, the legitimacy of a system is often referred to as people's trust in decision-making organs (Tyler & Jackson, 2013). In social psychology, a core finding of the literature on procedural justice is that authorities and institutions are more likely to be viewed as legitimate, once they exercise their power in a procedurally just manner (Tyler, 2006b, 2006a). Building on this central finding, we establish the following hypothesis:

H1: Procedural justice directly and positively influences the perceived legitimacy of a social scoring system.

According to the group value model of procedural justice (Lind & Tyler, 1988), central elements governing people's perceptions and behaviors are the procedures deployed by an authority. These procedures determine people's judgments of legitimacy, which, in turn, are important determinants of compliance with the authority. Therefore, to the extent that group procedures are perceived as just, legitimate judgments of the authority and compliance with it tend to increase (Lind & Earley, 1991). Both the suggestions of the group value model, as well as empirical findings (Moorman et al., 1998; Tyler, 2006b; Tyler & Fagan, 2006) propose that procedural justice is related to compliance intentions over the mediating variable perceived legitimacy. In the context of our study, these findings imply that perceptions of legitimacy directly and positively influence people's willingness to comply (Dickson et al., 2022). In addition, perceptions of legitimacy carry over the positive effects of procedural justice on the intention to comply. As a result, we hypothesize:

H2a: Legitimacy mediates the positive effect of procedural justice on the intention to comply with a social scoring system.

H2b: Legitimacy directly and positively influences the intention to comply with a social scoring system.

Outcome Favorability

Outcome favorability measures people's satisfaction with a specific outcome; hereby, "outcome" can refer to an (automated) decision (Martin & Waldman, 2022; Wang et al., 2020), or to a specific experience with a system (Mazerolle et al., 2013; Tyler & Fagan, 2006). The extent to which a person perceives an outcome as favorable is important for understanding how that person evaluates the decision-making organ; people who have a favorable experience are more likely to evaluate a system, or an algorithmic decision as procedurally just (Ambrose & Kulik, 1989; Wang et al., 2020). In these cases, they are also more likely to consider a decision-making organ legitimate (Kostka, 2019; Martin & Waldman, 2022). In the context of ADM systems, this observation is also referred to as *outcome favorability bias* (Wang et al., 2020). We can thus expect a direct and positive path from outcome favorability to perceptions of procedural justice. In addition, we expect that outcome favorability impacts perceptions of legitimacy both directly (Kostka, 2019; Waldman & Martin, 2022), as well as indirectly over the mediating variable procedural justice (Brockner, 2002). Importantly, the strength of these links likely depends on the level of perceived procedural justice. In the context of social scoring systems, in turn, perceptions of procedural justice may be impacted by the level of transparency, as outlined in the following subsection.

Providing Transparency in Scoring Systems

Many scholars argue that at first glance, it is intuitive to prefer transparent over non-transparent processes (Waldman & Martin, 2022), as transparency should, in principle, enhance the legitimacy of ADM systems (K. de Fine Licht & de Fine Licht, 2020; Tiarks, 2021). Algorithmic opacity makes it difficult to interrogate ADM systems and decreases the extent to which these systems can be held accountable (Colaner, 2022; Innerarity, 2021). However, full transparency can also undermine trust in decision-making organs (J. de Fine Licht et al., 2014; Grimmelikhuijsen, 2012); people might be disillusioned when they are confronted with the decision-making principles if these do not align with their imagination of fair decision-making procedures (K. de Fine Licht & de Fine Licht, 2020). Alternatively, providing transparency about underlying processes may lead people to discover subjective injustices (Kizilcec, 2016; Lee et al., 2019).

We hypothesize first, that transparency in social scoring systems increases perceptions of procedural justice as compared to cases in which decision-making principles are obscure. On the one hand, several studies conceptualize transparency as a sub-element of procedural justice (Lee et al., 2019; Tyler & Jackson, 2013). On the other hand, transparency delivers both an increased understanding of the evaluation mechanisms (standard of clarity), as well as an increased knowledge about which behavior is necessary to improve the score (outcome transparency). In addition, we expect that transparency enhances different aspects of perceived legitimacy, as it predicts people's trust in, for example, state organs (Alessandro et al., 2021) or decision-making systems (Kizilcec, 2016). In addition, a successful manipulation of procedural justice, possibly through providing more transparency, can lead to higher perceptions of legitimacy (Mazerolle et al., 2013). Reversely, removing transparency might lead people to question the accuracy of the evaluation method, and increased insecurity about whether the evaluation mechanism is accurate decreases the legitimacy of the system (Waldman & Martin, 2022). Therefore, we hypothesize:

H3: Transparency increases perceptions of procedural justice.

H4: Transparency increases perceptions of legitimacy.

As higher levels of transparency are expected to lead to an increase in procedural justice, we hypothesize that the relationship between the core perceptional variables – favorability, procedural justice, and legitimacy – is affected. Research has shown that the importance of outcome favorability in explaining perceptions of procedural justice and legitimacy decreases in situations in which deployed mechanisms are perceived as procedurally just (Brockner, 2002; Tyler & Fagan, 2006). That is, the stronger perceptions of procedural justice, the weaker the impact of outcome favorability on both the perceived legitimacy of a system, as well as on perceptions of procedural justice. In the context of scoring systems, these observations imply that people who are subjectively "disfavored" from the system may still perceive the system as legitimate, provided they evaluate the deployed mechanisms as procedurally just. In such instances, even these subjectively "disfavored" individuals might decide to comply with the system (Tyler, 1994). The increased perception of procedural justice in a transparent system likely undermines the importance of outcome favorability in explaining perceptions of legitimacy. That is, perceived legitimacy is not influenced by or dependent on a favorable experience with the system, once the evaluation mechanism of the system is made transparent. Further, the increased perception of procedural justice in a transparent system is expected to weaken the direct impact of favorability on perceptions of procedural justice. Lastly, we expect that the positive direct path from outcome favorability to procedural justice only exists once a scoring system is opaque, as perceptions of procedural justice are expected to be lower than under a transparent system. Therefore, we hypothesize:

H5a: Outcome favorability positively and indirectly influences perceptions of legitimacy over the mediating variable procedural justice only under a non-transparent scoring system.

H5b: Outcome favorability positively and directly influences perceptions of legitimacy only under a non-transparent scoring system.

H5c: Outcome favorability positively and directly influences perceptions of procedural justice only under a non-transparent scoring system.

Subjective Privacy Harms

Subjective privacy harms stem from the unwanted state of being observed, to which "unwelcome mental states" are attached (Calo, 2011). Subjective privacy harms may negatively impact people's willingness to adopt the behaviors that are desired by a system; related research has shown that an unwanted feeling of being monitored may crowd out any compliance behaviors that are above the expectations voiced by an authority (Niehoff & Moorman, 1993). Thus, in case an ADM system makes people feel observed we expect that the intention to comply is negatively and directly affected. In the digital age, algorithmic decision-making usually relies on an automated and large-scale collection of behavioral data. Therefore, concerns about data collection and processing likely also impact people's perceptions of ADM systems (Araujo et al., 2020); the perception of being involuntarily watched may constitute a significant privacy threat to individuals (Ohm, 2010), which might be negatively associated with the perceived justice of the procedures deployed by ADM systems (Araujo et al., 2020; Thurman et al., 2019). Therefore, we hypothesize:

H6a: If people feel surveilled, their intention to comply is directly and negatively impacted.

H6b: If people feel surveilled, perceptions of procedural justice are directly and negatively impacted.

The hypothesized model is illustrated in Figure 1. In the following, the scoring experiment and the survey are described. Subsequently, we present the analytical method to test our research model.

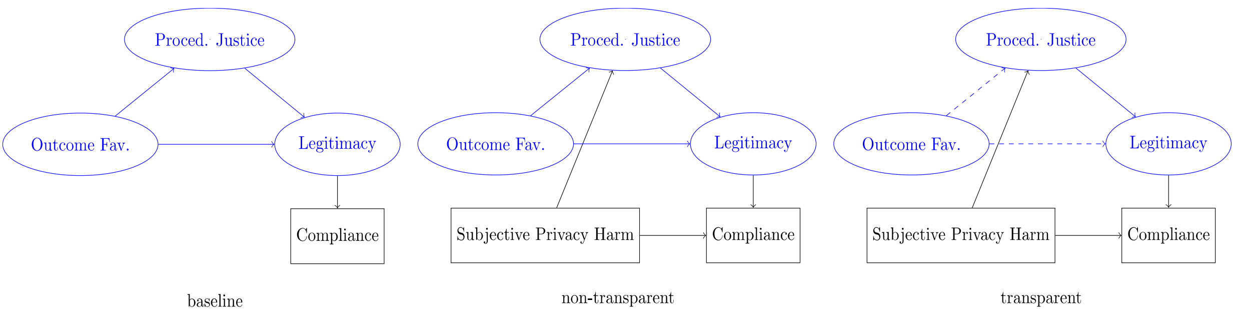


Figure 1. Proposed structural model

Method

Social Scoring Experiment

The social scoring experiment investigated people's behaviors and perceptions of an institutionalized intervention system, aiming at educating people on fair behaviors. The experiment was conducted in the ExperimenTUM lab of the Technical University of Munich in August and October 2022. The analysis contains data from 148 participants, most of whom were university students. 59.5% of the participants were aged between 17 and 24, 34.5% were aged between 25 and 30, and 4.1% were older than 30. 48.0% of the participants were female, 52.0% male. 57.4% were of White-Caucasian, 27.0% were of Asian-Pacific, and 4.7% of Hispanic origin. 4.7% indicated to be multiracial, and 6.1% did not reveal their ethnicity.

The experiment resembled a community game of trust (Duffy et al., 2013). A trust game was used building on the premise that trustworthiness is an instance of pro-social behavior (Letki 2006), mimicking people's fairness-related behavior towards each other. With this approach, we aimed at exemplarily replicating people's experiences of an ADM system, trying to educate people to a specific type of behavioral change, i.e., to more trustworthy and thus fairer behaviors. The experiment was conducted in three variants (between-subject design), which differed in the mechanism used to incentivize participants to fair behaviors. At the beginning of the experiment, participants were randomly matched to anonymous communities consisting of four people. In each community, the participants interacted with each other for an indefinite number of rounds. In each round, the four community members were randomly matched to pairs of two and interacted with each other in a trust game (Berg et al., 1995). A trust game is a sequential economic game between a so-called first-mover (trustor) and a second-mover (trustee). At the beginning of each interaction, the roles (first and second mover) are determined randomly. Both the first and second movers receive a monetary amount (also called *endowment*); in our case 10 monetary units (MUs). The first mover decides how much to send to the second mover. The amount sent is multiplied before the second mover receives it. In the present study, the amount sent was doubled. Then, the second mover decides how much of the amount received to send back. As such, the first mover's decision is a measure of *trust*. The second movers' decision is a measure of *trustworthiness*. After all participants of a community had finished their interactions in one round, a die was rolled to determine whether the experiment continued to another round. The continuation chance amounted to 95%. The participants were remunerated based on the MUs collected during the experiment. An overview of the session characteristics is given in Table 1.

	transparent	non-transparent	baseline
Number of participants	51	50	47
Number of sessions	4	4	4
Number of participants per session	13.75 (9.11)	15 (7.26)	12.5 (8.35)
Number of rounds	16.75 (2.75)	22 (9.35)	21.5 (3.42)
Remuneration in Euro	11.65 (1.84)	16.80 (5.93)	14.37 (3.34)
Duration in minutes	37.0 (8.8)	46.7 (15.3)	31.0 (7.2)

Table 1. Mean values and standard deviations (in brackets) of the sessions.

In each community, an institutional intervention system was introduced, which tried to incentivize participants to trustworthy (fair) behaviors. Fairness meant that the payoffs between the first and second mover from one interaction were equal (fairness principle). For example, if the first mover sent 10 MUs, the second mover received 20 MUs. With the initial endowment of 10 MUs, the second mover disposed of 30 MUs. Fair behavior meant that the second mover should send back 15 MUs. Participants were educated about the fairness principle at the beginning of the experiment. We conducted three treatments, one baseline intervention, and two scoring interventions. In the two scoring interventions, the institutional intervention used a scoring mechanism, which evaluated each person's fairness behaviors over the course of the experiment to a so-called *standing*. The standing ranged from A to E, based on a numerical score between 1000 and 0, where A was the highest, and E was the lowest standing. The two scoring treatments differed in terms of the transparency provided to participants regarding the scoring mechanism. Everyone started with a standing of A, and a score of 1000. If people did not behave fairly, the score was reduced proportionally to the transgression against the fairness principle. Thereby, the stronger the transgression, the more points were subtracted. The trustees' standings were updated after each interaction. If the score fell under 985 (970, 955, 940), the standing was downgraded to B (C, D, E), respectively. In case the subtraction led to a downgrade of the trustee's standing, the participant was informed about the downgrade and reminded to respect the fairness principle. Once it was reduced, the score could not be elevated again. The institution made it possible for the trustors, but not for the trustees, to request the standing of their interaction partners before deciding how much to send. As the standing informed about the interaction partner's past trustworthiness, the first mover could make an informed trust decision.

In the transparent treatment, people received information about the underlying ranges and an update about their score and standing in each round. In the non-transparent intervention, people knew their standing. They were not informed about their score nor about the subtraction mechanism. As a reference, a baseline treatment with a “normative” institutional system was run. Here, the standings were calculated in the background. In case the hypothetical standing was downgraded, the institution reminded the impacted participant to treat each other in a fair manner. Thus, the timing of the intervention was the same in all treatments.

Post-Experimental Survey

After the experiment, a survey was administered to collect people’s perceptions and their intention to comply with the institutional system. All items, except for the intention to comply, were asked on a 5-point Likert scale. People’s intention to comply was measured through a hypothetical scenario. Participants should suppose that they had to play the experiment again and under the same conditions. They were asked how much they would send back as a second mover if, in the first round of the experiment, they received 20 MUs from the first mover. The amount stated is an indicator of their intention to comply.

The survey measured outcome favorability via the extent to which participants liked their experience with the institution (f1)[1], and to which they liked how they were treated by others (f2). Two more questions were addressed, one relating to whether the presence of the institution was perceived as favorable (f3), and one asking how favorable the outcome (i.e., their final standing) was to the participants (f4). The scale for perceived legitimacy was measured via the extent to which the goals of the institution were normatively aligned with participants’ viewpoints (na1, na2), over participants’ perceived obligation to behave in line with the postulated behavior (ob1), and their perceived trust in the institution (tr1) (Jackson et al., 2023).

Perceived procedural justice was measured through the dignity and respect people felt they received (one question, d1) (Tyler & Jackson, 2013), and perceived benevolence (three questions) (Alessandro et al., 2021). The latter investigated whether people thought the institution acted in favor of the community (b1), whether the decisions were good for the community (b2), and whether the institution did what was necessary to help the community achieve fairness (b3). We asked further questions to assess the clarity of the standards (c1), which refers to how well people understand the rules used by a decision-maker to make decisions (Lee et al., 2019). We further evaluated whether participants felt capable of influencing the process of the institutional evaluation, both generally (pc1), and also such that they were satisfied with the outcome, i.e. their standing (pc2) (Lee et al., 2019). Lastly, outcome transparency was measured by asking people about their general fairness perceptions of the institutional evaluation mechanism (t1) and whether they knew how their behavior would impact their standing (t2) (Lee et al., 2019). Lastly, drawing on Calo’s definition of subjective privacy harm, as “individual … perception of unwanted observation” (Calo, 2011) we asked participants to which degree they felt controlled by the institution (ctr1). An overview of the constructs and scales is given in Table 5 in the online Appendix. The experiment and the survey were coded in oTree (Chen et al., 2016), which is a standard software for economic experiments. Our institution does not require ethics approval for experimental and questionnaire-based studies. We followed standard practices for ethical research, e.g., presenting detailed study procedures, obtaining consent, collecting anonymized data, and allowing participants to leave at any time. We added an attention check, and excluded those who did not pass the check (8 participants).

Structural Equation Modeling

Building on our theoretical review and empirical findings of prior studies, we evaluate our proposed research model, using multi-group structural equation modeling. In a structural equation model (SEM) a measurement model and a path model are estimated. The measurement model determines the factor loadings of the observed (or manifest) variables on the latent variables. In our research, the latent variables are perceived procedural justice, perceived legitimacy, and perceived outcome favorability. In the path model, the structural relationships between the latent variables are computed. It is also possible to include observed variables as predictors or outcomes in a path model (Mazerolle et al., 2013). Multi-group structural equation models allow for analyzing multi-group data. Here, a measurement model and path model are estimated *separately* for each group. In our case, the grouping variable is the experimental treatment. With regard to the measurement model, a prerequisite for drawing meaningful comparisons from a multi-group approach is some form of measurement invariance (Widaman, 1987). Configural invariance suggests that only the number of latent variables and their loading patterns are equal across

[1] The abbreviations indicate the identifiers of the survey questions listed in Table 5 in the online Appendix (https://www.cs.cit.tum.de/fileadmin/w00cfj/ct/papers/2023_ICIS_Appendix.pdf).

groups. Weak invariance suggests that the factor loadings of the manifest on the latent variables are equal across groups. Strong invariance suggests that both the factor loadings and the intercepts are equal across groups (Kline, 2016). As for the path model, multi-group structural equation modeling allows to investigate whether specific paths of the hypothesized structural model differ across groups. Hereby, several models with different restriction criteria are specified. Models in which a path is assumed to be invariant across groups are more restricted than those in which a path is assumed to differ across groups. Comparing the model fits of less restricted models to that of more restricted ones, the best-fitting structural multi-group model can be determined. As such, multi-group structural equation modeling constitutes a suitable approach to evaluate our structural model, as well as the impact of different transparency levels. It is important to note that structural models show a hypothesized structural relationship between latent and manifest variables. SEM cannot be used to establish causality between these variables (Bullock et al., 1994).

In the next section, two structural equation models are presented. The first model assesses all three interventions, namely the baseline, the non-transparent, and the transparent treatment. This allows us to compare people's perceptions of a scoring mechanism with their attitudes toward a normative intervention system. Here, we assess the structural paths between perceived legitimacy, procedural justice, and outcome favorability, as well as people's intentions to comply. The second SEM assesses the scoring interventions only. In the scoring treatments, people were asked a larger set of questions. Therefore, the second model allows for elaborating on the impact of transparency in greater detail. Here, we extend the model to investigate the impact of subjective privacy harms on people's perceptions and their willingness to comply. Both models were built following a stepwise approach; first, a measurement model between the three core latent variables Procedural Justice, Favorability, and Legitimacy was established. Hereby, all manifest variables that showed large cross-loadings, high residual error correlations, or low loadings on latent variables were removed, and the measurement model was re-specified accordingly (Kline, 2016). Second, a structural model was built to examine the relationships between these three core latent variables. Hereby, the model fit was iteratively improved, by comparing the goodness of fit of models with different restriction assumptions, using likelihood-ratio tests (LRT). Third, following the conceptual discussion in the previous section, further variables were included: in the first model, people's intention to comply (variable Compliance) was included. In the second model comprising only the scoring treatments, both people's intention to comply, as well as their subjective privacy harms (variable Subjective Privacy Harms) were included. As such, we can determine first, how subjective privacy harms impact people's perceptions as well as people's intention to comply (RQ1). Second, we can evaluate whether transparency moderates the influence of subjective privacy harms in explaining perceptions and compliance intentions (RQ2). We used the lavaan program in R (Rosseel, 2012). All estimations were done using the maximum-likelihood method (MLM) estimator, yielding robust standard errors.

Results

Model 1 – Reduced Model

Procedural Justice was measured in a 3-item-1-factor model, including participants' perceived benevolence (b1, b3) and received dignity (d1). Favorability was measured in a 3-item-1-factor model using two favorability questions (f1, f2), and allowing the dignity question (d1) to cross-load. Legitimacy was constructed over the items "normative alignment" (na1, na2), and "obligation to obey" (ob1). To determine the path model, we tested the core model as indicated by the blue variables in Figure 1. Weak measurement invariance was given. We found that the best-fitting structural model contained no path from Favorability to Legitimacy. The impact of Procedural Justice on Legitimacy was invariant and significant across groups. In contrast, the path from Favorability to Procedural Justice varied across groups; a significant impact was found in the baseline and non-transparent intervention, but no significant impact was found in the transparent intervention. Including Compliance as a manifest variable, we found a significant impact from Legitimacy to Compliance, which was invariant across groups. The model fit the data acceptably (CFI=1.00, robust RMSEA=0.00, SRMR=0.08, χ^2=502.533, *df*=84, *p*=<0.001) (Hu & Bentler, 1999). The effects are shown in Table 2. The LRT statistics are shown in Table 6 in the online Appendix.

Results for Model 1 – Reduced Model

We expected that perceiving institutional mechanisms as procedurally just is key for making people consider an institution legitimate (H1). Our model confirmed this relationship; in all treatments, perceived legitimacy was enhanced once the applied procedures were considered just. We further expected that perceptions of legitimacy are key to enhancing a willingness to comply (H2b). Moreover, perceptions of legitimacy were expected to carry over perceptions of justice to the willingness to comply (H2a). In all treatments, compliance was directly and positively influenced by perceptions of legitimacy, suggesting that a legitimate status, regardless of which kind of system, is important for evoking an intention to comply. The importance of a legitimate status for making people willing to comply is further emphasized by the mediating role of legitimacy; in all treatments, perceived legitimacy mediated the positive effects of perceived procedural justice on compliance intentions. Thus, both H2a and H2b are confirmed.

We expected that under a non-transparent scoring system, outcome favorability would contribute to making people perceive the system as just (H5c). In addition, we hypothesized that under a non-transparent scoring system, outcome favorability would explain perceptions of legitimacy, both directly (H5b), as well as indirectly (H5a). We referred to this as outcome favorability bias, as these effects are expected to be absent once transparency is introduced. Our model suggests that under a non-transparent scoring mechanism, a favorable outcome is a prerequisite for people to perceive the institutional operating principles as just. Against our expectations in H5b, we did not find a direct impact from outcome favorability to perceptions of legitimacy in the non-transparent system. However, in the non-transparent treatment, outcome favorability impacted perceived legitimacy in an *indirect* way over the mediating variable Procedural Justice; perceptions of procedural justice carried over the effect from a favorable experimental outcome to a high perception of perceived legitimacy. Once the scoring mechanism was made transparent, in contrast, the importance of favorable outcomes in explaining procedural justice and perceived legitimacy disappeared. As such, our model provides support for H5a and H5c. Our model further revealed that in the non-transparent scoring intervention, outcome favorability was important for indirectly explaining people's intention to comply. Those who received a favorable outcome were slightly more likely to intend to comply. Again, this effect is absent once the mechanism is made transparent. Interestingly, the importance of outcome favorability for explaining perceptions and the willingness to comply, which are only present under a non-transparent scoring system, also occur under a merely normative intervention system (Table 2). As such, we can determine that making a scoring mechanism transparent eliminates the outcome favorability bias (Wang et al., 2020) that is present under an opaque scoring system or a normative intervention system.

Baseline	Criterion	DE	IE	SE	CI-low	CI-high	p-value
Favorability	PJ	0.83***		0.05	0.74	0.93	0.00
	Legitimacy		0.76***	0.08	0.60	0.92	0.00
	Compliance		0.21*	0.08	0.06	0.37	0.01
PJ	Legitimacy	0.71***		0.09	0.53	0.90	0.00
	Compliance		0.26*	0.10	0.06	0.46	0.01
Legitimacy	Compliance	0.23***		0.08	0.07	0.39	0.00
Non-Transparent							
Favorability	PJ	0.77***		0.10	0.58	0.96	0.00
	Legitimacy		0.70**	0.11	0.49	0.92	0.00
	Compliance		0.20*	0.08	0.04	0.36	0.02
PJ	Legitimacy	0.91***		0.09	0.75	1.08	0.00
	Compliance		0.26*	0.10	0.06	0.46	0.01
Legitimacy	Compliance	0.28***		0.10	0.09	0.48	0.00
Transparent							
Favorability	PJ	0.26		0.18	-0.08	0.61	0.14
	Legitimacy		0.24	0.16	-0.08	0.56	0.14
	Compliance		0.07	0.06	-0.04	0.18	0.23
PJ	Legitimacy	0.86**		0.09	0.68	1.04	0.00
	Compliance		0.26*	0.10	0.06	0.46	0.01
Legitimacy	Compliance	0.17*		0.07	0.04	0.30	0.01

Table 2. Decomposition of Direct Effects (DE) and Indirect Effects (IE) for the Reduced Model (PJ: Procedural Justice)

The following section investigates the effects of transparency in the scoring interventions in more detail. We first test whether transparency resulted in a manipulation of perceptions of procedural justice and legitimacy. Then, we estimate the structural model for the scoring interventions, investigating the impact of subjective privacy harms on people's perceptions of the scoring system and their intention to comply.

Model 2 – Scoring Model

We hypothesized that transparency increases both perceptions of procedural justice (H3), as well as of legitimacy (H4). Perceptions of legitimacy are indeed significantly higher in the transparent (*M*=3.78, *SD*=0.72), as compared to the non-transparent treatment (*M*=3.50, *SD*=0.81, *p*=0.042; right-sided Mann-Whitney-U test). Additionally, perceptions of procedural justice are significantly higher in the transparent (*M*=4.14, *SD*=0.56), as compared to the non-transparent (*M*=3.77, *SD*=0.82, *p*=0.017) treatment. As such, we find support for both H3 and H4.

As for the measurement model, Procedural Justice was measured over five items, relating to benevolence (b1, b2), process control (pc1, pc2), and outcome transparency (tr1). Legitimacy was measured over four items – normative alignment (na1, na2), obligation to obey (ob1), as well as trust in the institution (tr1). Favorability was measured over two items (f1, f2). An overview of the correlations between the used items and Cronbach's alpha is shown in Table 4 in the online Appendix. Regarding the path model and similar to the previous model, no path from Favorability to Legitimacy was found. The path from Favorability to Procedural Justice was moderated through the transparency intervention, and highly significant in the non-transparent (β=0.81, *p*<0.001), but insignificant in the transparent treatment. In addition, our model suggests a variant path from Procedural Justice to Legitimacy. However, the estimates are very similar and significant across the treatments (β=0.96, *p*<0.001 in the non-transparent, and β=0.95, *p*<0.001 in the transparent intervention). In the scoring model, too, Compliance is directly predicted by Legitimacy. Hereby, the importance of Legitimacy for explaining Compliance is higher in the non-transparent (β=0.24, *p*<0.05) than in the transparent intervention (β=0.18, *p*<0.05). In addition, Compliance is impacted by Subjective Privacy Harms. This effect, however, only occurs in the transparent case; while there is no specific relationship between Subjective Privacy Harms and Compliance under a non-transparent mechanism, perceived subjective privacy harms significantly decrease people's willingness to comply under a transparent system (β=-0.35, *p*<0.001). Lastly, we found that in our specific experiment, Subjective Privacy Harms negatively influenced the variable Favorability. However, our model suggests that this negative impact only occurs in a transparent scoring system. The resulting effects are shown in Table 3. The resulting structural model with coefficients is found in Figure 2 (robust CFI=.955, robust RMSEA=.060, SRMR=.093, χ^2=156.077, *df*=134, *p*=0.093).

Non-Transparent	Criterion	DE	IE	SE	CI-low	CI-high	p-value
SPH	Favorability	-0.20		0.21	-0.60	0.21	0.34
	PJ		-0.16	0.17	-0.49	0.17	0.34
	Legitimacy		-0.15	0.16	-0.47	0.16	0.33
	Compliance	-0.04		0.17	-0.37	0.30	0.82
	Compliance		-0.04	0.04	-0.12	0.04	0.37
Favorability	PJ	0.81***		0.06	0.69	0.92	0.00
	Legitimacy		0.77***	0.07	0.63	0.91	0.00
	Compliance		0.42*	0.19	0.05	0.79	0.02
PJ	Legitimacy	0.96***		0.04	0.88	1.03	0.00
	Compliance		0.23*	0.10	0.03	0.43	0.03
Legitimacy	Compliance	0.24*		0.11	0.03	0.45	0.03
Transparent							
SPH	Favorability	-0.46*		0.18	-0.80	-0.11	0.01
	PJ		-0.08	0.07	-0.23	0.06	0.25
	Legitimacy		-0.08	0.07	-0.22	0.05	0.24
	Compliance	-0.35***		0.11	-0.56	-0.14	0.00

	Compliance		-0.02	0.02	-0.06	0.02	0.34
Favorability	PJ	0.19		0.15	-0.10	0.47	0.21
	Legitimacy		0.18	0.14	-0.10	0.45	0.21
	Compliance		0.28*	0.14	0.01	0.55	0.04
PJ	Legitimacy	0.95***		0.04	0.87	1.03	0.00
	Compliance		0.23*	0.10	0.03	0.42	0.03
Legitimacy	Compliance	0.18*		0.08	0.02	0.34	0.03

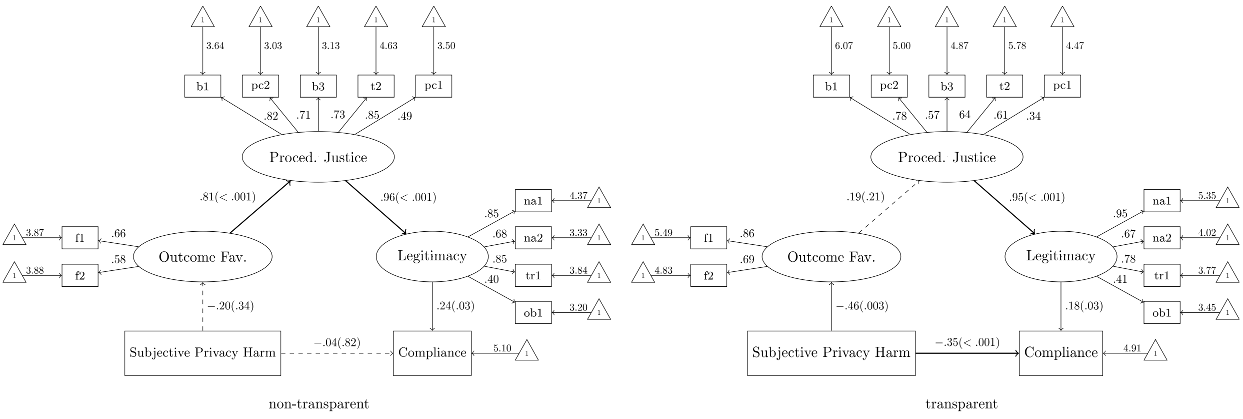
Table 3. Decomposition of Direct Effects (DE) and Indirect Effects (IE) for the Scoring Model (SPH: Subjective Privacy Harm, PJ: Procedural Justice)

To contextualize the analytical strength of the scoring model, we estimated the prospective power of the model in terms of detecting the outcome favorability bias. We used the package semPower (Moshagen, 2022) and followed the procedures in (Jobst et al., 2023). We estimated the model assuming factor loadings of 0.6 and a smallest effect size of interest of 0.3 for all paths except from Favorability to Procedural Justice. In addition, we assumed a difference in the direct path from Favorability to Procedural Justice between the groups of 0.2 (0.4). Given the sample size of 50 participants per group, the power to detect the outcome favorability bias for one variant path (Favorability to Procedural Justice) across groups amounts to 79% (82%).

Results for Model 2 – Scoring Model

As for the relationship between the core variables, the scoring model confirmed the outcome favorability bias; subjective outcome favorability was important in determining perceptions of justice (H5c), and legitimacy (H5b), as well as the intention to comply, only once the scoring mechanism was opaque. For the scoring interventions, we expected that subjective privacy harms would decrease people's perceptions of procedural justice (H6b). Against this expectation, subjective privacy harms did not impact perceptions of procedural justice. Rather, subjective privacy harms were important in determining people's outcome favorability, but only in the transparent scoring treatment. Although subjective privacy harms did not significantly differ between the transparent treatment (*M*=2.90, *SD*=1.03), and non-transparent treatment (*M*=2.74, *SD*=1.15, *p*=0.440; two-sided Mann-Whitney-U test), they substantially decreased people's satisfaction with the outcome in the transparent treatment (β=-0.46, p<0.05), but did not exert any impact in the non-transparent treatment.

We further hypothesized that the feeling of being surveilled under a scoring system would negatively impact people's willingness to comply with the behaviors postulated by the system (H6a). Indeed, the perception that people's privacy is invaded by a *transparent* institutional scoring system made people less likely to comply with institutional goals in the future. No such effects were found in the non-transparent case. As such, H6a is confirmed for only the transparent case. As for the impact of subjective privacy harms on people's perceptions and willingness to comply (RQ1), we can therefore determine that subjective privacy harms had a significant negative impact on outcome favorability *only* under a transparent mechanism. In addition, subjective privacy harms negatively impacted people's willingness to comply *only* under a transparent mechanism. Therefore, answering RQ2, transparency moderates the importance of subjective privacy harms on people's perceived outcome favorability, as well as on their intention to comply. The resulting paths (RQ1) and effects (RQ2) are illustrated in Figure 2.



non-transparent

transparent

Figure 2. Path estimates (p-values) of the scoring model.

Discussion and Concluding Remarks

Using structural equation modeling, we investigated the determinants of the perceived legitimacy of, as well as people's intention to comply with social scoring systems, aiming at educating people to fair behaviors. Adopting a system-level perspective, we examined social scoring systems with different levels of transparency regarding the scoring process. Our experiment showed that the provision of transparency successfully manipulated people's perceptions of procedural justice and legitimacy. Providing information about the underlying score calculation mechanism may represent a form of justification for the decisions made by the scoring system; justifications are considered a limited form of transparency, which potentially increase perceptions of legitimacy of ADM systems (K. de Fine Licht & de Fine Licht, 2020). Our results provide empirical support for this suggestion. The increased perceptions of procedural justice in a transparent scoring system determined people's perceptions of the scoring system following the principles of well-established justice theories (Brockner, 2002; Thibaut & Walker, 1975; Tyler, 2006b). Specifically, we found that when institutional decision-making processes of the scoring system were unclear, or when the intervention to encourage certain behaviors was only normative, people's perceptions of procedural justice were primarily determined by their overall favorable experiences with the system. In these cases, outcome favorability further strongly contributed to explaining people's perceptions of legitimacy, which is also referred to as "outcome favorability bias" (Wang et al., 2020). As such, our hypothesized theoretical model regarding the relationship between procedural justice, legitimacy, and people's intention to comply was nearly entirely validated. Our results, therefore, imply that procedural justice theories (Lind & Tyler, 1988; Thibaut & Walker, 1975) can be used to evaluate people's perceptions of modern ADM systems. Nevertheless, our findings also underscore that once these theories are used to study the implications of ADM systems, they need to be refined to incorporate aspects that are relevant in the digital age. Specifically, the data-collection practices of modern ADM systems may raise privacy-related concerns, such as imposing subjective privacy harms on individuals. Our work confirmed the findings of previous studies, suggesting that online privacy concerns are negatively associated with attitudes towards ADM systems (Araujo et al., 2020; Thurman et al., 2019). Yet, our expectations of *how* subjective privacy harms would impact people's perceptions and compliance intention did not entirely align with our hypothesized rationales. First, against our expectation, subjective privacy harms did not exert any impact on perceptions of procedural justice; however, they were negatively associated with outcome favorability, but only in a transparent system. Second, people's intention to comply was negatively impacted only in the transparent scoring system. These results imply that in the context of ADM systems, privacy and transparency may operate interdependently. To elaborate, transparency may lead to an increased awareness of the privacy-invading character of modern data collection processes, which, in turn, affects people's perceptions of ADM systems, as well as their behavioral intentions. Our results, therefore, suggest that further conceptual work is needed to understand the relationship between privacy and transparency, as well as their joint effect on people's judgments of ADM systems. Ashworth and Free, for example, have conceptualized privacy as an exchange framework between data collectors and individuals. They suggest that privacy-related practices of data controllers in the digital age affect not only perceptions of procedural justice, but also outcome favorability-related judgments. From a procedural justice-based perspective, subjective privacy harms arise when data controllers violate specific norms, leading individuals to experience a sense of disrespect. From an outcome favorability-based perspective, subjective privacy harms arise once the valuation of people's input is disproportional to the benefits people receive (Ashworth & Free, 2006). This conceptualization may explain why, in the context of our study, subjective privacy harms were negatively associated with outcome favorability. Following this conceptualization, participants in our study seemed to perceive a discrepancy between the feeling of being surveilled and the valuation of their monetary compensation. Our work, therefore, implies that the ongoing debate about the right level of transparency in ADM system may also consider privacy-related aspects (Ananny & Crawford, 2018; Burrell, 2016).

From a practical perspective, our results suggest first, that scoring systems that are characterized by opacity may give people a wrong sense of security regarding their online privacy, and may induce people to misjudge the use and value of their data. Second, with regard to the outcome favorability bias, Wang et al. have argued that "systems that evaluate algorithm fairness through user feedback must measure or incorporate an "outcome favorability" bias in their models" (Wang et al., 2020). However, our results open the possibility for another option: if system designers successfully achieve that decision subjects perceive the procedures deployed by ADM systems as just, the individual-level outcomes that people experience being subject to the system may not be important in determining whether they perceive the system as just or legitimate. High perceptions of justice would thus render the need to incorporate favorability biases in models that try to assess the fairness of a system less important.

However, further studies are needed to validate our findings and suggestions. With our work, we experimentally evaluated people's experiences with a social scoring system in interactions that are characterized by insecurity about whether strangers can be trusted. Our study was, therefore, limited to study one specific kind of incentive problem. In addition, monetary rewards could be retrieved from having a good score. To develop a comprehensive understanding of the impacts of providing transparency in social scoring systems, future studies may investigate the implications of using non-monetary rewards in scoring contexts, or of using scoring mechanisms to incentivize other instances of pro-social behaviors. Further, efforts to understand the implications of social scoring systems on groups of people that differ in socio-demographic or cultural characteristics could further enhance our understanding of social scoring.

Acknowledgments

We wish to thank the anonymous reviewers for their valuable feedback. We further thank Felix Fischer, Emmanuel Syrmoudis, and the facilitators of the ExperimenTUM lab for their support. We are grateful for funding support from the Bavarian Research Institute for Digital Transformation (bidt). Responsibility for the contents of this publication rests with the authors.

Bibliography

Alessandro, M., Lagomarsino, B. C., Scartascini, C., Streb, J., & Torrealday, J. (2021). Transparency and Trust in Government. Evidence from a Survey Experiment. *World Development*, 138. **https://doi.org/10.1016/j.worlddev.2020.105223**

Ambrose, M. L., & Kulik, C. T. (1989). The Influence of Social Comparisons on Perceptions of Procedural Fairness. *Journal of Business and Psychology*, 4(1), 129–138. **https://www.jstor.org/stable/25092220**

Ananny, M., & Crawford, K. (2018). Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability. *New Media & Society*, 20(3), 973–989. **https://doi.org/10.1177/1461444816676645**

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. (2020). In AI We Trust? Perceptions About Automated Decision-making by Artificial Intelligence. *AI & SOCIETY*, 35, 611–623. **https://doi.org/10.1007/s00146-019-00931-w**

Azeroual, A., Taher, Y., & Nsiri, B. (2020). Recidivism Forecasting: A Study on Process of Feature Selection. *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*. **https://doi.org/10.1145/3386723.3387848**

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10(1), 122–142. **https://doi.org/10.1006/game.1995.1027**

Brockner, J. (2002). Making Sense of Procedural Fairness: How High Procedural Fairness Can Reduce or Heighten the Influence of Outcome Favorability. *The Academy of Management Review*, 27(1), 58–76. **https://doi.org/10.2307/4134369**

Brockner, J., & Wiesenfeld, B. M. (1996). An Integrative Framework for Explaining Reactions to Decisions: Interactive Effects of Outcomes and Procedures. *Psychological Bulletin*, 120(2), 189–208. **https://doi.org/10.1037/0033-2909.120.2.189**

Büchi, M., Fosch-Villaronga, E., Lutz, C., Tamò-Larrieux, A., & Velidi, S. (2023). Making Sense of Algorithmic Profiling: User Perceptions on Facebook. *Information, Communication & Society*, 26(4), 809–825. **https://doi.org/10.1080/1369118X.2021.1989011**

Bullock, H. E., Harlow, L. L., & Mulaik, S. A. (1994). Causation Issues in Structural Equation Modeling Research. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(3), 253–267. **https://doi.org/10.1080/10705519409539977**

Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, 3(1). **https://doi.org/10.1177/2053951715622512**

Calo, R. M. (2011). The Boundaries of Privacy Harm. *Indiana Law Journal*, 86(3), 1132–1162.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97. **https://doi.org/10.1016/j.jbef.2015.12.001**

Citron, D., & Pasquale, F. (2014). The Scored Society: Due Process for Automated Predictions. *Washington Law Review*, 89, 1–33.

Colaner, N. (2022). Is Explainable Artificial Intelligence Intrinsically Valuable? *AI & SOCIETY*, 37, 231–238. **https://doi.org/10.1007/s00146-021-01184-2**

Cristianini, N., & Scantamburlo, T. (2020). On Social Machines for Algorithmic Regulation. *AI & Society*, 35, 645–662. **https://doi.org/10.1007/s00146-019-00917-8**

de Fine Licht, J., Naurin, D., Esaiasson, P., & Gilljam, M. (2014). When Does Transparency Generate Legitimacy? Experimenting on a Context-Bound Relationship. *Governance*, 27(1), 111–134. **https://doi.org/10.1111/gove.12021**

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial Intelligence, Transparency, and Public Decision-Making: Why Explanations Are Key When Trying to Produce Perceived Legitimacy. *AI & Society*, 35(4), 917–926. **https://doi.org/10.1007/s00146-020-00960-w**

Dickson, E. S., Gordon, S. C., & Huber, G. A. (2022). Identifying Legitimacy: Experimental Evidence on Compliance with Authority. *Science Advances*, 8(7). **https://doi.org/10.1126/sciadv.abj7377**

Drozdal, J., Weisz, J., Wang, D., Dass, G., Yao, B., Zhao, C., Muller, M., Ju, L., & Su, H. (2020). Trust in AutoML: Exploring Information Needs for Establishing Trust in Automated Machine Learning Systems. *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 297–307. **https://doi.org/10.1145/3377325.3377501**

Duffy, J., Xie, H., & Lee, Y.-J. (2013). Social Norms, Information, and Trust Among Strangers: Theory and Evidence. *Economic Theory*, 52(2), 669–708. **https://doi.org/10.1007/s00199-011-0659-x**

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding Explainability: Towards Social Transparency in AI Systems. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–19. **https://doi.org/10.1145/3411764.3445188**

Grimmelikhuijsen, S. (2012). Linking Transparency, Knowledge and Citizen Trust in Government: An Experiment. *International Review of Administrative Sciences*, 78, 50–73. **https://doi.org/10.1177/0020852311429667**

Hardin, R. (1993). The Street-Level Epistemology of Trust. *Politics & Society*, 21(4), 505–529. **https://doi.org/10.1177/0032329293021004006**

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. **https://doi.org/10.1080/10705519909540118**

Innerarity, D. (2021). Making the Black Box Society Transparent. *AI & SOCIETY*, 36, 975–981. **https://doi.org/10.1007/s00146-020-01130-8**

Jackson, J., Bradford, B., Giacomantonio, C., & Mugford, R. (2023). Developing Core National Indicators of Public Attitudes towards the Police in Canada. *Policing and Society*, 33(3), 276–295. **https://doi.org/10.1080/10439463.2022.2102757**

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 624–635. **https://doi.org/10.1145/3442188.3445923**

Jobst, L., Bader, M., & Moshagen, M. (2023). A Tutorial on Assessing Statistical Power and Determining Sample Size for Structural Equation Models. *Psychological Methods*, 28(1), 207–221. **https://doi.org/10.1037/met0000423**

Kizilcec, R. F. (2016). How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2390–2395. **https://doi.org/10.1145/2858036.2858402**

Kline, R. B. (2016). Principles and Practice of Structural Equation Modeling (4th ed.). The Guilford Press. **https://doi.org/10.1017/CBO9780511614934**

Kostka, G. (2019). China's Social Credit Systems and Public Opinion: Explaining High Levels of Approval. *New Media & Society*, 21(7), 1565–1593. **https://doi.org/10.1177/1461444819826402**

Lee, M. K. (2018). Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management. *Big Data & Society*, 5. **https://doi.org/10.1177/2053951718756684**

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., & Kusbit, D. (2019). Procedural Justice in Algorithmic Fairness: Leveraging Transparency and Outcome Control for Fair Algorithmic Mediation. *Proceedings of the ACM on Human-Computer Interaction, 3 (CSCW)*. **https://doi.org/10.1145/3359284**

Letki, N. (2006). Investigating the Roots of Civic Morality: Trust, Social Capital, and Institutional Performance. *Political Behavior*, 28(4), 305–325. **https://doi.org/10.1007/s11109-006-9013-6**

Li, L., Lassiter, T., Oh, J., & Lee, M. K. (2021). Algorithmic Hiring in Practice: Recruiter and HR Professional's Perspectives on AI Use in Hiring. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society,* 166–176. **https://doi.org/10.1145/3461702.3462531**

Lind, E., & Earley, P. (1991). Some Thoughts on Self and Group Interests: A Parallel-Processor Model. *The Annual Meeting of the Academy of Management*, Miami, 251–259.

Lind, E., & Tyler, T. (1988). The Social Psychology of Procedural Justice. In *Contemporary Sociology* (Bd. 18). **https://doi.org/10.2307/2073346**

Martin, K., & Waldman, A. (2022). Are Algorithmic Decisions Legitimate? The Effect of Process and Outcomes on Perceptions of Legitimacy of AI Decisions. *Journal of Business Ethics*, 183. **https://doi.org/10.1007/s10551-021-05032-7**

Mazerolle, L., Antrobus, E., Bennett, S., & Tyler, T. (2013). Shaping Citizen Perceptions of Police Legitimacy: A Randomized Field Trial of Procedural Justice. *Criminology*, 51(1), 33–63. **https://doi.org/10.1111/j.1745-9125.2012.00289.x**

Moorman, R., Blakely, G., & Niehoff, B. (1998). Does Perceived Organizational Support Mediate the Relationship Between Procedural Justice and Organizational Citizenship Behavior? *Academy of Management Journal*, 41, 351–357. **https://doi.org/10.2307/256913**

Moshagen, M. (2022). semPower: Power Analyses for SEM. R package version 1.2.0. **https://cloud.r-project.org/web/packages/semPower/semPower.pdf**

Niehoff, B. P., & Moorman, R. H. (1993). Justice as a Mediator of the Relationship between Methods of Monitoring and Organizational Citizenship Behavior. *The Academy of Management Journal*, 36(3), 527–556. **https://doi.org/10.2307/256913**

Njuguna, R., & Sowon, K. (2021). Poster: A Scoping Review of Alternative Credit Scoring Literature. *ACM SIGCAS Conference on Computing and Sustainable Societies*, 437–444. **https://doi.org/10.1145/3460112.3471972**

Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review,* 57, 1701–1777. **https://www.uclalawreview.org/pdf/57-6-3.pdf**

O'Reilly, T. (2013). Open Data and Algorithmic Regulation. In B. Goldstein & L. Dyson (Eds.), Beyond Transparency: Open Data and the Future of Civic Innovation, 21, 289–300.

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as Mechanisms for Supporting Algorithmic Transparency. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. **https://doi.org/10.1145/3173574.3173677**

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. **https://doi.org/10.18637/jss.v048.i02**

State Council. (2014). Planning Outline for the Construction of a Social Credit System (2014-2020). *State Council of China*.

Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *The Academy of Management Review*, 20(3), 571–610. **https://doi.org/10.2307/258788**

Sunshine, J., & Tyler, T. (2003). The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing. *Law & Society Review*, 37(3), 513–548.

Thibaut, J. W., & Walker, L. (1975). Procedural Justice: A Psychological Analysis. L. Erlbaum Associates. **https://api.semanticscholar.org/CorpusID:1464011**

Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My Friends, Editors, Algorithms, and I. *Digital Journalism*, 7(4), 447–469. **https://doi.org/10.1080/21670811.2018.1493936**

Tiarks, E. (2021). The Impact of Algorithms on Legitimacy in Sentencing. *Journal of Law, Technology and Trust*, 2(1). **https://doi.org/10.19164/jltt.v2i1.1151**

Tyler, T. (1994). Governing amid Diversity: The Effect of Fair Decisionmaking Procedures on the Legitimacy of Government. *Law & Society Review*, 28(4), 809–831. **https://doi.org/10.2307/3053998**

Tyler, T. (2006a). Psychological Perspectives on Legitimacy and Legitimation. *Annual Review of Psychology*, 57, 375–400. **https://doi.org/10.1146/annurev.psych.57.102904.190038**

Tyler, T. (2006b). Why People Obey the Law. *Princeton University Press*. **http://www.jstor.org/stable/j.ctv1j66769**

Tyler, T., & Fagan, J. (2006). Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in Their Communities? *Ohio State Journal of Criminal Law*, 6, 231–276. **https://scholarship.law.columbia.edu/faculty_scholarship/414**

Tyler, T., & Jackson, J. (2013). Popular Legitimacy and the Exercise of Legal Authority: Motivating Compliance, Cooperation and Engagement. *Psychology Public Policy and Law*, 20(1), 78–95. **https://doi.org/10.1037/a0034514**

Uslaner, E. (2002). The Moral Foundations of Trust. *Cambridge University Press*. **https://doi.org/10.1017/CBO9780511614934**

Vereschak, O., Bailly, G., & Caramiaux, B. (2021). How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2). **https://doi.org/10.1145/3476068**

Waldman, A., & Martin, K. (2022). Governing Algorithmic Decisions: The Role of Decision Importance and Governance on Perceived Legitimacy of Algorithmic Decisions. *Big Data & Society*, 9(1). **https://doi.org/10.1177/20539517221100449**

Wang, R., Harper, F. M., & Zhu, H. (2020). Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. **https://doi.org/10.1145/3313831.3376813**

Widaman, K. F. (1987). Journal of Educational Statistics, 12 (3), 208-313. **https://doi.org/10.1145/3313831.3376813**

Wolfangel, E. (2022). Nein, Bayern bereitet keine Überwachung chinesischer Art vor. Zeit. **https://www.zeit.de/digital/datenschutz/2022-07/oeko-token-bayern-belohnungssystem-social-scoring**

Yeung, K. (2018). Algorithmic Regulation: A Critical Interrogation. *Regulation & Governance*, 12(4), 505–523. **https://doi.org/10.1111/rego.12158**

Zarsky, T. (2015). Understanding Discrimination in the Scored Society. *Washington Law Review*, 89(4), 1375–1412.