

# On the Factors that Shape Perceptions Towards Decision-Makers in Algorithmic Decision-Making

## 1 INTRODUCTION

In an effort to ensure that normatively *fair* algorithmic decision-making processes are perceived as such, an increasing number of studies are now looking into the elements that impact fairness perceptions in algorithmic decision-making (e.g., [4, 25, 27, 28]). Despite this recent attention, there is still a lack of nuanced understanding about what drives perceptions towards a core procedural factor that determines how algorithmic decision-making processes unfold: the *decision-makers* themselves. This, in turn, makes it challenging to understand *whether* and *why* fairness perceptions towards algorithmic decision-making processes might be moderated by users' perceptions towards the decision-maker.

To explore the properties that decision subjects value of decision-makers in *algorithmic* decision-making processes and factors that contribute to those properties, we conducted an initial exploratory interview-based study with 21 participants for a holiday rental scenario<sup>1</sup>. Through a *reflexive thematic analysis* [5, 6] and by combining an inductive and deductive orientation to data, we found that our participants' perceptions towards the profile and configuration of decision-makers were, overall, well aligned with the model of organizational (perceived) trustworthiness (i.e., ability, benevolence, integrity) suggested by Mayer et al. [22]. Based on those findings, we formulate the following research questions that will guide our work:

- **RQ1:** Do factors related to the profile, and configuration of decision-makers shape perceptions of ability, benevolence, and integrity towards decision-makers in algorithmic decision-making?
- **RQ2:** Do perceptions of ability, benevolence, and integrity towards decision-makers predict overall fairness perceptions towards algorithmic decision-making processes?

## 2 STUDY SETUP AND HYPOTHESES

**Study type.** Randomized controlled trial; between-subjects design with  $(2 \times 2 \times 2 = 8)$  groups.

**Context.** Holiday rental scenario.

**Hypotheses.**

- **Hypotheses related to RQ1: Factors affecting Perceived Ability, Benevolence, Integrity.**

**Hypothesis 1a ( $H_{1a}$ ).** A human decision-maker that uses an algorithmic system to augment their capabilities is perceived as more able than an algorithmic decision-maker.

**Rationale.** Previous work suggests that algorithmic decision-making processes with a fully algorithmic decision-maker are perceived to be efficient and objective [17, 28]. However, these

are also perceived to be less adaptable than humans [13]. Participants in our exploratory study highlighted that a hybrid decision-making setup (i.e., human decision-maker with an algorithmic system) benefits from the *ability* of the algorithmic system to efficiently and accurately process data, while enabling the human to exercise discretion. We, therefore, hypothesize that a human using an algorithmic system will be perceived as more able than a fully algorithmic decision-maker.

**Hypothesis 1b ( $H_{1b}$ ).** A human decision-maker that uses an algorithmic system to augment their capabilities is perceived as more benevolent than an algorithmic decision-maker.

**Rationale.** Previous work, through qualitative findings, suggests that algorithmic decision-makers are considered impersonal and dehumanizing [4]. Problematic aspects of an algorithmic decision-maker include its inability to account for the unique individual circumstances of decision subjects, and to adapt the decision-making to their needs and preferences [18, 28]. In our exploratory study, our participants highlighted, that a decision-making process that is augmented by an algorithmic system, but, where the final decision is made a human, can show empathy and consideration towards the decision subject, i.e., is able to be more *benevolent*. We, therefore, hypothesize that a human decision-maker using an algorithmic system will be perceived to be more benevolent than an algorithmic decision-maker.

**Hypothesis 1c ( $H_{1c}$ ).** The perceived integrity of a decision-maker is higher when it concerns rule-based models than when it concerns a probabilistic model.

**Rationale.** Binns et al. [4], through their qualitative findings, suggested that decision subjects consider statistical inferences unacceptable as a basis for algorithmic decision-making. Similarly, some participants of our exploratory study claimed that generalization should not be acceptable as a decision basis. Participants, in contrast, were asking for a clear indication of the rules that they were violating and that motivated a first warning. Even if Wang et al. [27] did not find any effect of the model type on decision subjects' *overall* fairness perceptions, we hypothesize that relying on rule-based models will contribute to higher perceptions of *integrity* compared to probabilistic models.

**Hypothesis 1d ( $H_{1d}$ ).** The perceived integrity of a decision-maker is higher when the data used for decision-making comes from publicly available databases rather than non-publicly available data sources.

**Rationale.** Anik and Bunt [1] suggested that information about data sources used for training a model allow users to judge the trustworthiness of a system and to assess its fairness. Participants in our exploratory study highlighted the role of the data

<sup>1</sup>We chose an algorithmic system suggested by the municipality of Amsterdam for detecting illegal holiday rentals as a use case. <https://algorithmeregister.amsterdam.nl/en/illegal-holiday-rental-housing-risk/> (last accessed 28.11.2023)

type used as part of the decision-making in relation to the *proportionality* of the means that a decision-maker uses to inform their decision. Participants suggested that it is acceptable to use publicly available data, while accessing data that might invade the privacy of decision subjects (i.e., non-publicly available data) was not considered acceptable. We, therefore, hypothesize that using non-publicly available data for decision-making will negatively impact decision subjects' perceptions of integrity towards the decision-maker as compared to using publicly available data.

- **Hypothesis related to RQ2: Effect of Perceived Ability, Benevolence, Integrity on Perceptions of Fairness.**

**Hypothesis 2a (H<sub>2a</sub>). Perceived ability relates positively to perceptions of fairness.**

*Rationale.* Previous literature in human decision-making did not find *ability* to be a significant predictor for fairness perceptions [7]. As opposed to these findings, we hypothesize that a difference in context might play a role. Colquitt and Rodell [7] studied the relationship between perceived ability and perceptions of fairness by recruiting alumni from a university and capturing their perceptions towards their immediate managers. For this context, the authors argued that more able managers might create more outcome differentiation in their units, which the alumni might not always benefit from, and therefore, might not perceive as fair. As opposed to this context, we hypothesize that in a context where citizens might benefit from higher levels of ability in the decision-maker (e.g., by ensuring that, thanks to detecting illegal holiday rentals, the societal issue of not having enough long-term rentals available is ameliorated) perceived ability will relate positively to fairness perceptions.

**Hypothesis 2b (H<sub>2b</sub>). Perceived benevolence relates positively to perceptions of fairness.**

*Rationale.* Prior literature in human decision-making found that for *benevolence* and integrity the relationships between perceived trustworthiness and fairness perceptions are reciprocal, with both influencing one another [7]. Similarly, we hypothesize that in *algorithmic* decision-making, *benevolence* will relate positively to fairness perceptions.

**Hypothesis 2c (H<sub>2c</sub>). Perceived integrity relates positively to perceptions of fairness.**

*Rationale.* Literature in human decision-making has shown that perceptions of *integrity* affect all dimensions of fairness perceptions (i.e., distributive, procedural, informational, interpersonal) [7]. We hypothesize that for *algorithmic* decision-making processes, there will also be a positive relation between perceived integrity and perceptions of fairness.

### 3 METHOD

#### 3.1 Variables

**3.1.1 Independent variables.** Tables 1 and 2 show how each independent variable is displayed in practice. AI refers to Artificial Intelligence, and we use it to refer to the algorithmic system.

- **Profile (categorical, between-subjects).** Each participant will be randomly assigned to one of two configurations (Table 2):

- (1) AI-Human (hybrid)<sup>2</sup>. Participants will be presented a scenario where an AI is used as a screening tool that informs the decision of the human civil servant to consider the reported property an illegal holiday rental. The human civil servant evaluates the output of the system and, based on their own judgement [24], decides whether to send a first warning to the property owner.
- (2) Only AI (fully algorithmic). Participants will be presented a scenario where a fully automated decision-making involves an AI evaluating the reported property and, based on that evaluation, determining whether there is an illegal holiday rental in that address. Based on the output of the AI system, a warning letter is sent to the property owner.

- **Model type (categorical, between-subjects).** Each participant will be randomly assigned to one of two configurations:

- (1) Probabilistic. Participants will be presented a scenario where the AI system calculates the probability of the reported address to be an illegal holiday rental based on a set of parameters. Each parameter is followed by a different number of (+) signs to indicate that some of those parameters had a more prominent impact on the final probability [4, 8].
- (2) Rule-based. Participants will be presented a scenario where the AI system evaluates whether the reported address meets relevant conditions that might indicate the property is being illegally rented as a holiday rental.

The parameters that the probabilistic and rule-based models consider depend on the type of data that the AI system retrieves. If publicly available data is retrieved, we present participants with a few of the parameters that the original system suggested by the municipality of Amsterdam relies on for calculating a probability<sup>3</sup>. We made sure that none of those parameters are protected by law [3] (e.g., gender). If data that is not publicly available is retrieved, we present participants with parameters related to the flow of people accessing the building, as suggested by a couple of the participants in our exploratory study. During the pilot test, we made sure that these parameters were considered believable.

- **Data provenance (categorical, between-subjects).** Each participant will be randomly assigned to one of two configurations:

- (1) Publicly available databases. Participants will be presented a scenario where the AI system has access to and retrieves information available in the public registry. This configuration is informed by the working of the original system suggested by the municipality of Amsterdam.
- (2) Non publicly available data sources. Participants will be presented a scenario where the AI system has access to and retrieves the camera footage from the doorbell in the building. If the doorbell has no camera, it accesses the footage from the nearest street camera. This configuration is informed

<sup>2</sup>The study was pilot tested with 12 experts in human-computer interaction from our institution. During that pilot test we checked the effectiveness of the manipulations, the feasibility of the presented scenarios [2], the layout, wording and potential biases that we might trigger [9].

<sup>3</sup>See <https://algorithmeregister.amsterdam.nl/en/illegal-holiday-rental-housing-risk/> (last accessed 28.11.2023)

Your city has limited living space; both for citizens and visitors. If a citizen wants to rent out their home on Airbnb to tourists, they need to meet certain requirements. They must also request a license to the municipality. Not everyone adheres to those conditions. The municipality sometimes receives reports that a home has been rented out without meeting the requirements. Until now, a human civil servant would manually investigate the report and find evidence that would help determine whether the reported property was being illegally rented.

Given the shortage of long-term rentals in your city, the municipality has decided to increase its efforts to identify citizens who do not meet the requirements to rent their homes on Airbnb. For this reason the municipality of your city has adopted an Artificial Intelligence system to accelerate the identification of these illegal rentals. With the new system, when a report is filed, the Artificial Intelligence system has access to **[Data provenance]**.

Based on that data, the Artificial Intelligence system **[Model type]** **[Profile]** and it is the first time that this address is reported, a first warning is sent to request the owner to stop renting the property illegally. After this first warning, the owner might face penalties if they fail to adhere to the vacation rental policy.

*[We present the diagram of the workflow. We provide an example to illustrate the workflow in practice.]*

A few hours ago, a report was filed to complain about a potential case of an illegal holiday rental in 25 Green Hill Street. After retrieving **[Data provenance (...)]**, the evaluation of the Artificial Intelligence is the following:  
**[Model type] ∩ [Data provenance]**

Since **[Model type (...)]**, **[Profile (...)]**. The letter includes a first warning and a request to stop renting the property illegally. It also includes information on how to **[Profile (...2)]** to ask any questions the 25 Green Hill Street owner might have.

**Table 1: Scenario presented to participants.**

by examples given by a couple of participants in our exploratory study on what constitutes a decision-making process that relies on disproportional means to enforce policy. During the pilot test, we made sure that this configuration was considered believable. We also contrasted the proportionality of requesting access to the camera footage with other examples of fraud detection where decision-makers request access to personal information (e.g., number of toothbrushes in a property to know whether all people registered in an address are in reality living there)<sup>4</sup>, so that participants do not consider the suggested scenario as too far away from a potential future [2].

3.1.2 *Dependent variables.* See Appendix A for the measurement instruments.

- **Perceived ability** (*continuous*). Measured by the average score on the six items suggested by Höddinghaus et al. [13]. These items adapt the original items suggested by Mayer and Davis [21] for ability to capture two relevant facets in algorithmic decision-making: data processing capacity and adaptability to changing conditions.
- **Perceived benevolence** (*continuous*). Measured by the average score on the five items suggested by Mayer and Davis [21].
- **Perceived integrity** (*continuous*). Measured by the average score on the six items suggested by Mayer and Davis [21].
- **Perceived fairness** (*continuous*). Measured by a one-item construct on a 7-point Likert scale, following previous work [16, 17, 28].

<sup>4</sup>See the example on how fraud is detected in student financial aid <https://verfassungsblog.de/a-scandal-on-ai-in-administration-again/> (last accessed 28.11.2023)

3.1.3 *Descriptive and exploratory measurements.* See appendix B for the measurement instruments.

- **Age group** (*categorical*). Age group that participants belong to. Participants will choose one of the six categorical options.
- **Level of education** (*categorical*). Highest level of education that participants have completed. Participants will choose one of the six categorical options.
- **Lessee of short-term rentals** (*categorical*). Participants will answer whether they have experience renting out their property as a short-term rental. Our exploratory study only included participants with experience renting their properties out. Even if the confirmatory study seeks to capture perceptions of the wider public and presents the scenario in the third person, we seek to understand whether having experience as a lessee of short-term rentals and, therefore, having a personal stake in the topic [17], has an impact on perceptions towards decision-makers.
- **AI literacy** (*continuous*). AI literacy has been shown to impact fairness perceptions in algorithmic decision-making [25, 28]. We will capture AI literacy through an adapted version of the scale used by Schoeffer et al. [25]. Measured by the average score on the four suggested items.
- **Affinity to technology** (*continuous*). Affinity to technology has been shown to affect the perceived ability of algorithmic systems [16]. We will capture it through an adapted version of the scale used by Franke et al. [11], following previous work [16, 28]. Measured by the average score on the four suggested items.
- **Personal experience with decision-makers of illegal short-term rentals** (*continuous*). Experience and familiarity with a specific decision-maker profile (algorithmic or non algorithmic)

Parameters	Conditions	Descriptions
Profile	AI-Human	<i>the human civil servant in charge examines the evaluation of the Artificial Intelligence. If, based on the civil servants' judgement, there are clear signs that indicate an illegal holiday rental in this address, (...) the human civil servant in charge has examined the evaluation of the Artificial Intelligence. Based on the civil servant' judgement, there are indeed clear signs that indicate an illegal holiday rental in this address. The human civil servant has, therefore, send a letter to the property owner of 25 Green Hill Street. (...2) contact the human civil servant in charge,</i>
	Only AI	<i>- (...) the evaluation of the Artificial Intelligence system has led to a letter to be sent to the property owner of 25 Green Hill Street. (...2) interact with the Artificial Intelligence system</i>
Model type	Probabilistic	<i>calculates the probability of a property being illegally rented on the reported address. If the probability is high, (...) the probability of this property being illegally rented is high,</i>
	Rule-based	<i>evaluates through a rule-based system whether the reported address meets the conditions of illegal holiday rental. If relevant conditions are met that indicate an illegal holiday rental in this property, (...) relevant conditions are met that indicate an illegal holiday rental in this property,</i>
Data provenance	Publicly available databases	<i>the public registry, where it retrieves information about prior illegal housing cases, about the building and about the identity and housing rights of the residents. (...) information from the public registry</i>
	Non publicly available data sources	<i>the camera footage of the doorbell in the building. If the doorbell has no camera, then it accesses the footage of the nearest street camera. Thanks to this footage, the AI identifies the flow of people accessing the building. (...) footage from the cameras</i>
Model $\cap$ Data	Probabilistic $\cap$ Public	<p><i>"The property in 25 Green Hill Street has a high probability probability of being an illegal holiday rental. According to the information in the public registry, the following factors determine the high probability:</i></p> <ul style="list-style-type: none"> <li><i>- Street code +++</i></li> <li><i>- Anonymous reporter +++</i></li> <li><i>- Number of rooms ++</i></li> <li><i>- Date of residence in the address +"</i></li> </ul> <p><i>(+) means that this factors contributed to getting a high probability. The more (+) signs, the bigger the impact of that factor on getting a high probability.</i></p>
	Probabilistic $\cap$ Non-public	<p><i>"The property in 25 Green Hill Street has a high probability probability of being an illegal holiday rental. According to the information obtained from the camera in the last month, the following factors determine the high probability:</i></p> <ul style="list-style-type: none"> <li><i>- Total number of suitcases detected entering the building +++</i></li> <li><i>- Total number of non-regular residents entering the building +++</i></li> <li><i>- Affluence of people during weekends and holidays ++</i></li> <li><i>- Frequency of access of people during working hours +"</i></li> </ul> <p><i>(+) means that this factors contributed to getting a high probability. The more (+) signs, the bigger the impact of that factor on getting a high probability.</i></p>
	Rule-based $\cap$ Public	<p><i>"The property in 25 Green Hill Street meets the conditions for being flagged as an illegal holiday rental. According to the information in the public registry, the following conditions were met:</i></p> <ul style="list-style-type: none"> <li><i>- The property is located in a highly touristic area of the city</i></li> <li><i>- The complaint is not anonymous, it comes from the neighbour nextdoor</i></li> <li><i>- The property has more than 2 rooms</i></li> <li><i>- The property owner is not registered in this address and has several other properties"</i></li> </ul>
	Rule-based $\cap$ Non-public	<p><i>"The property in 25 Green Hill Street meets the conditions for being flagged as an illegal holiday rental. According to the information obtained from the camera, the following conditions were met in the last month:</i></p> <ul style="list-style-type: none"> <li><i>- Total number of suitcases detected entering the building &gt; 15</i></li> <li><i>- Total number of non-regular residents entering the building &gt; 50</i></li> <li><i>- Affluence of people during weekends and holidays &gt; 5 people entering the building on average every 30 minutes during the day</i></li> <li><i>- Affluence of people during working hours &gt; 3 people entering the building on average every hour"</i></li> </ul> <p><i>These conditions apply to this particular building based on its size and factors such as the presence of other Airbnb-s in the building.</i></p>

**Table 2: Experimental design.**

has been shown to lead to preferences towards that decision-maker [15]. We will capture participants' personal experience with algorithmic systems or humans making decisions about illegal holiday rentals through an adapted version of the scale used by Kramer et al. [15]. Measured by the average score on the two suggested items.

- **Personal experience with public administration** (*continuous*). From our exploratory study, we observed that our participants' previous experiences with the public administration affected their perceptions towards the suggested scenarios. We will, therefore, employ an adapted version of the scale used by Kramer et al. [15]. Measured by the average score on the two suggested items.
- **Affinity to short-term rental policy** (*continuous*). From our exploratory study, we observed that our participants' perceptions towards the adequacy of the policy itself affected their perceptions towards the suggested scenarios. We will, therefore, measure affinity to the policy through a one-item construct on a 7-point Likert scale, following previous work [23].
- **Perceived task complexity** (*continuous*). Previous work has shown that task complexity affects preferences towards human or algorithmic decision-makers [17, 24]. We will, therefore, measure perceived task complexity through a one-item construct on a 7-point Likert scale, following previous work [20, 28].

### 3.2 Planned sample

We plan to recruit at least 205 participants for data collection purposes. We calculated our planned sample by using the software *G\*Power* [10], for a between-subjects ANOVA (*Fixed effects, special, main effects and interactions*) within the *F tests* family. We calculated the sample size by setting the default effect size 0.25, a significance threshold of  $\alpha = 0.05/7 = 0.007$  since we will test several hypothesis on the same data, a desired power of 0.8, with 8 groups and the respective degrees of freedom.

We will recruit our participants on *Prolific*<sup>5</sup> where we will share the link to our study with them. The study will be conducted on *Qualtrics*<sup>6</sup>. All our participants will be at least 18 years old and will participate in the study only once. We will screen participants to ensure that they are located in a country in the Global North (fairness perceptions towards algorithmic decision-making have been shown to vary depending on whether participants belong to the Global North or South [14]) and that they are proficient in English. Our study includes two attention checks. Participants who do not pass both attention checks will be discarded from the data analysis.

### 3.3 Procedure

**Step 1.** Participants are shown information about the study purpose and the way the data will be managed. Participants need to accept the informed consent reviewed by our institution to be able to proceed to the study. Participants then respond to questions related to their age, education level, experience as lessees of short term rentals, AI literacy, affinity to technology, personal experience

with short-term rentals, personal experience dealing with public administration and affinity to the policy regulating short-term rentals.

**Step 2.** Participants are shown a brief paragraph with information about the policy of their municipality in matters of short-term rentals. The introductory paragraph also mentions how illegal holiday rentals have been dealt with until now (i.e., by having a human civil servant manually inspect every reported address) so that participants can have an internal frame of reference they can evaluate the new configuration against. Participants are then introduced to the decision of the municipality to introduce an Artificial Intelligence system to accelerate the detection of illegal holiday rentals. Depending on which of the  $(2 \times 2 \times 2 = 8)$  between-subject scenarios participants get randomly assigned to, they will read about a workflow where a fully automated or a hybrid algorithmic decision-making process is put in place. Participants will also get to know whether the system relies on a probabilistic or rule-based model and whether it operates on publicly available or non-publicly available data. Participants will then be shown a graphical representation of the workflow to facilitate comprehension<sup>7</sup>.

**Step 3.** Participants are then shown an example of how the workflow looks in practice. To this end, we use an example where the property in 25 Green Hill Street has been detected as an illegal holiday rental. We deliberately frame the scenario in third-person to avoid *outcome favourability bias* [19, 27] among our participants. The purpose of this third step is to illustrate what we mean by terms like *probabilistic* or *rule-based* model. The decision to do so is based on the observations from our exploratory study, where participants, especially with lower AI literacy levels, would not understand what jargon-heavy terms would entail in practice until they saw an example. Participants then answer to the first attention check, where they are asked to select the purpose of the municipality when implementing this system.

**Step 4.** Participants are asked to evaluate the perceived ability, benevolence and integrity towards the decision-maker. Each set of items is preceded by a button where participants can select to be reminded about the workflow and another button where participants can select to be reminded about the example<sup>8</sup>. After each set of items, participants are asked to answer the corresponding open-ended question. The second attention check is located between the questionnaire about perceived ability and perceived benevolence. In this second attention check, participants are asked to select the technology that the municipality relies on to identify illegal holiday rentals. Participants are finally asked to evaluate their fairness perceptions towards the algorithmic decision-making process.

### 3.4 Analysis plan

We will analyze the hypotheses we formulated in Section 2 in four separate statistical analyses. First, to test **H<sub>1a</sub>** we will use a one-way ANOVA with *profile* as between-subject factor and *perceived ability*

<sup>7</sup>The graphical representations for each scenario were designed based on the feedback we got from the experts in the pilot study. We made sure that the graphic clearly shows the direction of the workflow and we avoided any image that could make participants anthropomorphize the AI system or link human-like intelligence traits to it (e.g., by avoiding to represent the AI through a brain and a human-looking robot).

<sup>8</sup>The experts in our pilot study recommended this option to avoid participants to forget the specifics about the presented scenario.

<sup>5</sup><https://www.prolific.com/>

<sup>6</sup><https://www.qualtrics.com/>

as dependent variable. To test  $H_{1b}$ , we will use a one-way ANOVA with *profile* as between-subject factor and *perceived benevolence* as dependent variable. To test  $H_{1c}$  and  $H_{1d}$ , we will conduct a multi-way ANOVA with *model type* and *data provenance* as between-subject factors and *perceived integrity* as dependent variable. To test  $H_{2a-c}$  we will conduct a multiple linear regression analysis with *perceived ability*, *benevolence*, and *integrity* as independent variables and *fairness perception* as dependent variable.

In case of deviations from the assumptions required for the parametrics tests mentioned above, we will use their non-parametric equivalent; in particular, Kruskal-Wallis test and non-parametric regression. Since we are testing 7 hypotheses on the same data, we apply a Bonferroni correction to our significance threshold, reducing it to  $\frac{0.05}{7} = 0.007$ .

We may conduct posthoc tests to analyze pairwise differences, mediation analyses and exploratory factor analyses to better understand our results.

For the open-ended questions, we will use *thematic analysis* to analyze the qualitative data.

## 4 ADDITIONAL COMMENTS

As of submitting this preregistration, data collection has not yet begun.

### A MEASUREMENT OF DEPENDENT VARIABLES

A. Items to measure *perceived ability*. Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

- (1) **[Decision-maker]**<sup>9</sup> has the competence to include all necessary information for making decisions about illegal holiday rentals.
- (2) **[Decision-maker]** is able to process all data necessary for making decisions about illegal holiday rentals.
- (3) **[Decision-maker]** is able to consider all necessary data when making decisions about illegal holiday rentals.
- (4) **[Decision-maker]** is capable of flexibly considering different circumstances when making decisions about illegal holiday rentals.
- (5) **[Decision-maker]** has the competence to adapt its decision to different circumstances.
- (6) **[Decision-maker]** is able to react flexibly to circumstances in the decision-making process.

B. Items to measure *perceived benevolence*. Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

- (1) **[Decision-maker]** will<sup>10</sup> take care of the welfare of the owner of 25 Green Hill Street.
- (2) **[Decision-maker]** will consider the needs and desires of the owner of 25 Green Hill Street.
- (3) **[Decision-maker]** will act on the best interest if the owner of 25 Green Hill Street.
- (4) **[Decision-maker]** will look out what is important for the owner of 25 Green Hill Street.
- (5) **[Decision-maker]** will go out of its way to help the owner of 25 Green Hill Street.

C. Items to measure *perceived integrity*. Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

- (1) **[Decision-maker]** acts with a strong sense of justice.
- (2) **[Decision-maker]** acts in an honest way.
- (3) **[Decision-maker]** is fair when identifying illegal holiday rentals.
- (4) The behaviours and decisions coming out of **[Decision-maker]** are not very consistent (r).
- (5) I like the values and purposes behind having a **[Decision-maker]** for identifying illegal holiday rentals.
- (6) Sound principles guide the behaviour of **[Decision-maker]**.

C. Item to measure *perceived fairness*. Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

- (1) Overall the decision-making process for identifying illegal holiday rentals set up by the municipality is fair.

### B MEASUREMENT OF DEMOGRAPHICS AND DESCRIPTIVE VARIABLES

A. Questionnaire for determining *age range*.

**What is your age range?**

- A1: 0-18
- A2: 19-25
- A3: 26-35
- A4: 36-50
- A5: 50-80
- A6: 80+

B. Questionnaire for determining *level of education*.

**What is the highest level of school that you have completed or the highest degree you have received?**

- A1: High school incomplete or less.
- A2: High school graduate or GED (includes technical / vocational training that does not award college credit)
- A3: Some college (some community college, associate's degree).

<sup>9</sup>**[Decision-maker]** is either "The Artificial Intelligence system" or "The human civil servant (by) using the Artificial Intelligence system and their own judgment [24]" depending on the condition that each participant gets.

<sup>10</sup>Following Gulati et al. [12] we use the future verb tense for *benevolence*. Benevolence is a relational facet of perceived trustworthiness [22] that evolves over time [26]. Solberg et al. [26] suggest that the effect of perceived ability and integrity are more salient at the beginning of the "relationship" between a decision-maker and a decision subject, whereas benevolence might evolve over time, as decision-maker and decision subject repeatedly interact. In our scenario, we open up the possibility for the 25 Green Hill Street owner to interact with the decision-maker and, therefore, encourage participants to evaluate how benevolent the decision-maker might be in that *future* interaction.

- A4: Four year college degree / bachelor's degree
- A5: Some postgraduate or professional schooling, no post-graduate degree
- A6: Postgraduate or professional degree, including master's, doctorate, medical or law degree

C. Items to determine experience as *lessee of short-term rentals*. Assessed as a yes/no question.

- (1) I have rented my house out for short-term rentals (for example, Airbnb) and I had a license for it.
- (2) I have rented my house out for short-term rentals and I did not have a license for it.

D. Items to measure *AI literacy*. Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

- (1) I have a good knowledge in the field of *artificial intelligence*.
- (2) My current employment includes working with *artificial intelligence*.
- (3) I am confident interacting with *artificial intelligence*.
- (4) I understand what the term *artificial intelligence* means.

E. Items to measure *affinity to technology*. Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

- (1) I like to occupy myself in greater details with technical systems (systems that include some technology: computing systems, electronic gadgets, mechanisms)
- (2) I like testing functions of new technical systems.
- (3) It is enough for me that a technical system works; I don't care about how or why (r)<sup>11</sup>.
- (4) It is enough for me to know the basic functions of a technical system (r).

F. Items to measure *personal experience with short-term rentals*. Assessed in a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

- (1) I am aware of human civil servants identifying illegal holiday rentals.
- (2) I am aware of artificial intelligence systems detecting illegal holiday rentals.

G. Items to measure *personal experience with public administration*. Assessed in a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

- (1) I have a good experience dealing with the human civil servants in the public administration.

H. Item to measure *affinity to short-term rental policy*. Assessed in a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

- (1) It is acceptable that the municipality enforces a policy to identify and penalize short-term rentals like Airbnb(s) that are not officially registered.

I. Item to measure *perceived task complexity*. Assessed in a seven-point Likert scale (1 = very low in complexity, 7 = very high in complexity).

- (1) How complex do you think it is to identify illegal holiday rentals?

J. Open-ended questions.

- (1) Do you think **[Decision-maker]** is capable of correctly identifying illegal holiday rentals? Why?
- (2) Do you think **[Decision-maker]** will treat the renter in 25 Green Hill Street with kindness? Why?
- (3) Do you think it is right that the municipality relies on **[Decision-maker]** for the decision-making process? Why?

## REFERENCES

- [1] Ariful Islam Anik and Andrea Bunt. 2021. Data-Centric Explanations: Explaining Training Data of Machine Learning Systems to Promote Transparency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. <https://doi.org/10.1145/3411764.3445736>
- [2] Jeffrey Bardzell, Shaowen Bardzell, and Erik Stolterman. 2014. Reading critical designs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1951–1960. <https://doi.org/10.1145/2556288.2557137>
- [3] Uladzislau Belavusau and Kristin Henrard. 2019. A Bird's Eye View on EU Anti-Discrimination Law: The Impact of the 2000 Equality Directives. *German Law Journal* 20, 05 (7 2019), 614–636. <https://doi.org/10.1017/glj.2019.53>
- [4] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage'; Perceptions of Justice in Algorithmic Decisions. (1 2018). <https://doi.org/10.1145/3173574.3173951>
- [5] Victoria Clarke and Virginia Braun. 2013. *Successful qualitative research: A practical guide for beginners*. Sage publications Ltd. 1–400 pages.
- [6] Victoria Clarke and Virginia Braun. 2021. *Thematic analysis: a practical guide*. SAGE Publications Ltd.
- [7] Jason A. Colquitt and Jessica B. Rodell. 2011. Justice, Trust, and Trustworthiness: A Longitudinal Analysis Integrating Three Theoretical Perspectives. *Academy of Management Journal* 54, 6 (12 2011), 1183–1206. <https://doi.org/10.5465/amj.2007.0572>
- [8] Jonathan Dodge, Q. Vera Liao, Yunfeng Zhang, Rachel K. E. Bellamy, and Casey Dugan. 2019. Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment. (1 2019). <https://doi.org/10.1145/3301275.3302310>
- [9] Tim Draws, Alisa Rieger, Oana Inel, Ujwal Gadiraju, and Nava Tintarev. 2021. A Checklist to Combat Cognitive Biases in Crowdsourcing. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 9, 1 (10 2021), 48–59. <https://ojs.aaai.org/index.php/HCOMP/article/view/18939>
- [10] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (5 2007), 175–91. <https://doi.org/10.3758/bf03193146>
- [11] Thomas Franke, Christiane Attig, and Daniel Wessel. 2019. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction* 35, 6 (4 2019), 456–467. <https://doi.org/10.1080/10447318.2018.1456150>
- [12] Siddharth Gulati, Sonia Sousa, and David Lamas. 2019. Design, development and evaluation of a human-computer trust scale. *Behaviour & Information Technology* 38, 10 (10 2019), 1004–1015. <https://doi.org/10.1080/0144929X.2019.1656779>
- [13] Miriam Höddinghaus, Dominik Sondern, and Guido Hertel. 2021. The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior* 116 (3 2021), 106635. <https://doi.org/10.1016/j.chb.2020.106635>
- [14] Shivani Kapania, Oliver Siy, Gabe Clapper, Azhagu Meena SP, and Nithya Sambasivan. 2022. "Because AI is 100% right and safe": User Attitudes and Sources of AI Authority in India. In *CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–18. <https://doi.org/10.1145/3491102.3517533>
- [15] Max F. Kramer, Jana Schaich Borg, Vincent Conitzer, and Walter Sinnott-Armstrong. 2018. When Do People Want AI to Make Decisions?. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA, 204–209. <https://doi.org/10.1145/3278721.3278752>
- [16] Markus Langer, Tim Hunsicker, Tina Feldkamp, Cornelius J. König, and Nina Grgić-Hlača. 2022. "Look! It's a Computer Program! It's an Algorithm! It's AI!": Does Terminology Affect Human Perceptions and Evaluations of Algorithmic Decision-Making Systems?. In *CHI Conference on Human Factors in Computing*

<sup>11</sup>Reverse coded

- Systems. ACM, New York, NY, USA, 1–28. <https://doi.org/10.1145/3491102.3517527>
- [17] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (1 2018). <https://doi.org/10.1177/2053951718756684>
- [18] Chiara Longoni, Andrea Bonezzi, and Carey K Morewedge. 2019. Resistance to Medical Artificial Intelligence. *Journal of Consumer Research* 46, 4 (12 2019), 629–650. <https://doi.org/10.1093/jcr/ucz013>
- [19] Henrietta Lyons, Tim Miller, and Eduardo Velloso. 2023. Algorithmic Decisions, Desire for Control, and the Preference for Human Review over Algorithmic Review. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 764–774. <https://doi.org/10.1145/3593013.3594041>
- [20] Henrietta Lyons, Senuri Wijenayake, Tim Miller, and Eduardo Velloso. 2022. What’s the Appeal? Perceptions of Review Processes for Algorithmic Decisions. In *CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–15. <https://doi.org/10.1145/3491102.3517606>
- [21] Roger C. Mayer and James H. Davis. 1999. The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology* 84, 1 (2 1999), 123–136. <https://doi.org/10.1037/0021-9010.84.1.123>
- [22] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An Integrative Model of Organizational Trust. *The Academy of Management Review* 20, 3 (7 1995), 709. <https://doi.org/10.2307/258792>
- [23] Ryan Merrill and Nicole Sintov. 2016. An Affinity-to-Commons Model of Public Support For Environmental Energy Policy. *Energy Policy* 99 (12 2016), 88–99. <https://doi.org/10.1016/j.enpol.2016.09.048>
- [24] Rosanna Nagtegaal. 2021. The impact of using algorithms for managerial decisions on public employees’ procedural justice. *Government Information Quarterly* 38, 1 (1 2021), 101536. <https://doi.org/10.1016/j.giq.2020.101536>
- [25] Jakob Schoeffer, Niklas Kuehl, and Yvette Machowski. 2022. “There Is Not Enough Information”: On the Effects of Explanations on Perceptions of Informational Fairness and Trustworthiness in Automated Decision-Making. (5 2022). <https://doi.org/10.1145/3531146.3533218>
- [26] Elizabeth Solberg, Magnhild Kaarstad, Maren H. Rø Eitheim, Rossella Bisio, Kine Reegård, and Marten Bloch. 2022. A Conceptual Model of Trust, Perceived Risk, and Reliance on AI Decision Aids. *Group & Organization Management* 47, 2 (4 2022), 187–222. <https://doi.org/10.1177/10596011221081238>
- [27] Ruotong Wang, F. Maxwell Harper, and Haiyi Zhu. 2020. Factors Influencing Perceived Fairness in Algorithmic Decision-Making. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376813>
- [28] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev, and Alessandro Bozzon. 2023. Disentangling Fairness Perceptions in Algorithmic Decision-Making: the Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–21. <https://doi.org/10.1145/3544548.3581161>