# What's the Appeal? Perceptions of Review Processes for Algorithmic Decisions

Henrietta Lyons
hlyons@student.unimelb.edu.au
School of Computing and Information Systems
The University of Melbourne
Melbourne, Victoria, Australia

Senuri Wijenayake
senuri.wijenayake@sydney.edu.au
School of Architecture, Design and Planning
The University of Sydney
Sydney, New South Wales, Australia

Tim Miller
tmiller@unimelb.edu.au
School of Computing and Information Systems
The University of Melbourne
Melbourne, Victoria, Australia

Eduardo Velloso
eduardo.velloso@unimelb.edu.au
School of Computing and Information Systems
The University of Melbourne
Melbourne, Victoria, Australia

## ABSTRACT

If you were significantly impacted by an algorithmic decision, how would you want the decision to be reviewed? In this study, we explore perceptions of review processes for algorithmic decisions that differ across three dimensions: the reviewer, how the review is conducted, and how long the review takes. Using a choice-based conjoint analysis we find that people *prefer* review processes that provide for human review, the ability to participate in the review process, and a timely outcome. Using a survey, we find that people also see human review that provides for participation to be the *fairest* review process. Our qualitative analysis indicates that the fairest review process provides the greatest likelihood of a favourable outcome, an opportunity for the decision subject and their situation to be fully and accurately understood, human involvement, and dignity. These findings have implications for the design of contestation procedures and also the design of algorithmic decision-making processes.

## CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

## KEYWORDS

algorithmic decision-making; contestability; reviewability; algorithmic fairness, accountability, and transparency

**ACM Reference Format:**
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 (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3491102.3517606

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
*CHI '22, April 29-May 5, 2022, New Orleans, LA, USA*
© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3517606

## 1 INTRODUCTION

From grading students [25] to predicting the likelihood of recidivism [5], algorithms are increasingly being used to make decisions that have a significant impact on people's lives. Numerous documents proposing guidelines for the "ethical" design, development, and use of algorithmic decision-making systems highlight the importance of being able to challenge or 'contest' high-consequence algorithmic decisions [37, 65].[1] The ability to contest significant decisions, whether made by human or algorithm, is important: At a practical level, contestation is a mechanism for error correction [75, 80] and for holding decision-makers to account [19]. Contestation is also a tool that can help promote fairness and justice in decision-making [40], protect individual rights [7], and support notions of human dignity and autonomy [4, 39]. Even if people do not get the outcome they desire, being able to appeal a decision can provide a sense of closure. However, there is limited guidance around how appeal processes for algorithmic decisions could or should be designed [14, 37, 39]. There are numerous examples of poorly designed appeal processes, where people impacted by algorithmic decisions are unable to contest the decision. This can be due to a lack of explanation or information being provided about the decision, there being no information about how to appeal, no response to appeals, or because there is no appeal process available [2, 35, 62, 69, 79]. In this work, we explore perceptions of different styles of review for algorithmic decisions and use these insights to consider design implications for algorithmic contestation processes.

The design of an appeal process will not only determine whether a person will be able to effectively contest the decision, but can lend weight to the perceived fairness of a decision-making process [48, 77], the legitimacy of the decision [40], and a person's trust in the decision-making entity (noting that trust is an important aspect of human-AI interaction [36]). Yet, designing appeal processes is complex, with a number of elements to consider such as how people can access the appeal process, who will review the decision on appeal, and how a review of the initial decision will be conducted [55]. There is no one-size-fits-all approach—indeed, appeal mechanisms for human decision-making range from informal complaints to highly structured legal appeals. In this work, we focus on one step

[1] In this paper, we use the terms 'contest', 'challenge', and 'appeal' interchangeably.

of the appeal process: the review of the initial algorithmic decision. Our aim is to explore perceptions of different review processes that can be used when contesting algorithmic decisions. Specifically, we ask the following:

***RQ1:*** *What type of review process do people prefer when contesting an algorithmic decision?*

***RQ2:*** *How do different review processes for algorithmic decisions impact perceptions of fairness?*

***RQ3:*** *How do personal characteristics (such as age, gender, and attitudes towards algorithmic decision-making) impact perceptions of reviewers of algorithmic decisions?*

Drawing from research on perceptions of algorithmic decision-making, procedural justice literature investigating perceptions of fairness of human decision-making, and work exploring contesting algorithmic decision-making (e.g. [84]), we designed six base review processes that differ on two dimensions: who the *reviewer* is (human or algorithm) and the *style of review* (a review of the functioning of the algorithmic decision-making system, a new decision based on the original information provided, or a new decision that incorporates any new information and objections that the decision subject wishes to provide). We explored perceptions of these review processes using two approaches. First, with an algorithmic loan decision as our context, we used a choice-based conjoint analysis to force a preference choice between 12 different pairs of review processes. The review processes differed over three attributes; in addition to the reviewer and the type of review, we also explored the time the review takes (7 days, 30 days, or 60 days). In the second part of our study, we asked participants to rate the fairness of the six base review processes and to explain their rating. We also collected information about participants' attitudes towards, and familiarity with, algorithmic decision-making.

We found that participants preferred reviews conducted by humans in a timely manner that allowed the decision subject to provide new information and objections in relation to the initial decision. Interestingly, participants placed more weight on the time taken for the review and the type of review than who the reviewer was when they chose their preferred review process. In line with the results of the choice-based conjoint analysis, we found that review processes with human reviewers were perceived as more fair than appeal processes with algorithmic reviewers, as were review processes that allowed the decision subject more 'voice' (the ability to object and provide new information) [50, 77]. The belief that humans would make better loan decisions than algorithms was associated with higher perceived fairness of human reviewers, and the belief that algorithms should make important decisions that impact people's lives was associated with higher perceptions of fairness of algorithmic reviewers. There was also an association between fairness ratings of the two reviewers. The results of our qualitative analysis indicate that the fairest review process is one that provides the greatest likelihood of a favourable outcome, an opportunity for the decision subject and their situation to be fully and accurately understood, human involvement, and dignity.

This research makes three contributions. First, we provide empirical evidence that when designing review processes, attention should be given to the type of review and how long that review will take, in addition to who conducts the review. Second, we extend the literature exploring perceptions of algorithmic decision-making to the context of contestation, and we provide evidence that attitudes towards algorithmic decision-making impact perceptions of algorithmic reviewers. Third, we find parallels between procedural justice literature and perceptions of algorithmic decision-making, which suggests that findings from research focusing on the procedural fairness of human decision-making can be used to inform the design of algorithmic decision-making processes.

# 2 RELATED WORK

## 2.1 Contesting algorithmic decisions

There is a difference between designing appeal processes and designing for 'contestability' [61, 80]. Kluttz and Mulligan [41] define contestability as a set of *"mechanisms for users to understand, construct, shape and challenge model predictions"*. So, while contestability as a design goal encompasses the ability to contest a decision, it goes further than this by envisioning a system that can be interacted with and influenced by a decision subject during decision-making. In work developing a machine learning system that evaluates the delivery of psychotherapy, Hirsch et al. [33] outline the importance of designing systems for contestability to allow those impacted by an assessment decision the chance to correct errors and record their disagreement with the system. Almada [4] advocates for "contestability by design" where the ability to contest is considered throughout the design, development, and implementation of a decision-making algorithm. This can be achieved through the incorporation of participatory design, explanations for decisions, and interface design that enables people impacted by algorithmic decisions to challenge the decision [4]. In one of the few works with an empirical emphasis, Vaccaro et al. [81] ran a series of workshops to explore participants' views on designing for contestability in relation to automated content moderation, uncovering three key themes: (1) the need for moderation to be representative of users; (2) the need for strong two-way communication between the users and the platform; and (3) the need for compassion by platforms when moderating content.

Much of the academic work that considers designing appeal processes for algorithmic decisions is conceptual in nature and relates to the interpretation of Article 22 in the European Union's *General Data Protection Regulation* (GDPR), which sets out a right to contest decisions that have been made using solely automated processing (see e.g. [4, 8, 72, 84]). Article 22 has been interpreted as requiring a post decision review process. For example, Wachter et al. [84] suggest four different ways to review an algorithmic decision that differ according to who the reviewer is (human or algorithm) and how a review could be conducted. Beyond the GDPR, based on an analysis of submissions made in response to Australia's proposed AI ethics principles [65], Lyons et al. [55] found that the ability to contest was conceptualised as a post decision process, requiring a number of steps including the provision of an explanation, a way for decision subjects to elect to contest the decision, and a review process. Vaccaro et al. [80] explored the impact of appeal processes on perceptions of algorithmic content moderation decisions: they designed three types of appeal that differed according to who the reviewer was (human or algorithm) and how to review would be carried out (a review of the initial decision with consideration of a written appeal by the decision subject or a re-evaluation of activity history). They found that none of the appeal conditions impacted participants' perceptions of fairness, accountability, transparency or control of algorithmic decisions compared to the "no appeal" condition.

In this work, we aim to understand what type of process people prefer for reviewing algorithmic decisions. This empirical work contributes to, and complements, the above literature, by exploring perceptions of processes that could be used to review algorithmic decisions.

## 2.2 Designing 'fair' decision-making processes

Perceptions of human decision-making and whether decisions are viewed as 'just' or 'fair' have been studied across social psychology [77], organisational psychology [16, 48], and the law [51]. Initially, these studies focused on 'distributive justice', which concerns the way outcomes are allocated and whether they satisfy rules of equity, equality, and need [1, 17]. In 1975, Thibaut and Walker [77] conducted a series of studies on people's perceptions of dispute resolution processes. They found that people perceived processes where they had 'process control'—the ability to choose what evidence will be provided and time to state their case—to be fairer than those without [77]. Work focusing on the perceived fairness of decision-making procedures is commonly referred to as 'procedural justice'. Distributive and procedural justice are two dimensions of justice that have repeatedly been shown to impact perceptions of decisions. In this paper, we focus on procedural justice.

There is no agreed-upon procedural justice framework. A number of different theoretical approaches have been proposed since Thibaut and Walker's seminal studies (e.g. [48, 52]). One of the most replicated findings is that having process control or 'voice'—the ability to provide information and input relevant to the decision—in the decision-making process positively impacts perceptions of fairness [16, 50, 77]. The procedural justice literature largely focuses on the fairness of an initial decision. In contrast, an appeal is a *secondary* decision-making process, occuring after an initial decision has been made. Appeals are often studied in relation to their impact on fairness perceptions of an initial decision or the organisation making that decision. For example, the availability of appeal processes has been found to positively impact fairness perceptions of performance appraisals [27], trust in management [3], satisfaction with dispute resolution processes [76], and job satisfaction [3].

There are also studies that focus on different types of appeal processes. For example, using archival records and a field survey, Conlon [18] showed that for appealing a parking violation, an in-person appeal resulted in an increased perception of voice compared to a written appeal, but this increase in voice did not translate to an increased perception of procedural justice. This is surprising given that increased voice in a procedure has been linked to higher perceptions of justice. Conlon [18] suggested that because people were able to choose between the types of appeal, they had been given the opportunity to be heard either in writing or verbally, and as such had participated in the decision-making process and perceived it as fair regardless of how they made their appeal.

Recently, researchers in human-computer interaction have drawn from procedural justice literature to explore how the design of algorithmic decision-making processes impacts perceptions of fairness [9, 29, 30, 47, 80, 83]. Generally, this research shows that justice principles that apply to human decision-making are also relevant for algorithmic decision-making. For example, Binns et al. [9] found that perceptions across four dimensions of justice (distributive justice, procedural justice, interactional justice, and informational justice) were correlated, as they are in human decision-making. In contrast to findings in procedural justice literature, Vaccaro et al. [80] found that the ability to appeal an automated content moderation decision had no impact on ratings of fairness, accountability, trustworthiness, or feelings of control in relation to automated decision-making. Vaccaro et al. suggest that this result could be attributed to the perceived illegitimacy of social platforms in general, but may also be due to the design of the appeal mechanisms. Indeed, past research proposes two elements to voice: (1) a person should be able to have their say; and (2) a person should feel as though their input was heard and appropriately considered [49]. Vaccaro et al.'s design provides for the first element, but not the second.

We draw from, and contribute to, the procedural justice literature by exploring whether allowing for 'voice' in review processes for algorithmic decision-making has similar results to the inclusion of voice in human decision-making contexts.

## 2.3 Perceptions of algorithmic decision-making

People's views of algorithms are measured in a variety of ways, for example via preferences [42], general attitudes [67], and perceptions of fairness [6, 46, 56, 85]. The findings of this research are mixed. Some work supports the notion of *'algorithmic aversion'*—negative views of algorithms in comparison to humans [22, 38]. For example, Dietvorst et al. [22] found that even when algorithms were shown to perform better than humans, when people saw that the algorithm could make mistakes they chose to rely on forecasts made by humans who made the same mistakes as the algorithm. Other research supports the concept of *'algorithmic appreciation'*—where people prefer the judgement of algorithms to humans [6, 38, 53]. For example, Logg et al. [53] found that lay people tend to rely more on advice if it comes from an algorithm than if it comes from a human.

These differences in findings can be attributed to a variety of factors, including the type of decision being made, the decision context, and personal characteristics. For example, perceptions of algorithmic decision-making have been shown to be influenced by views of what algorithms are capable of, people's knowledge of algorithms, and people's familiarity with algorithmic decision-making in specific contexts. Kramer et al. [42] and Castelo et al. [13] found that when participants were familiar with the use of algorithmic decision-making in a particular context they preferred the use of algorithms in those contexts. Nagtegaal [63] found that algorithms were perceived as more just when the task was low in complexity (e.g. determining reimbursement for travel expenses) and humans were seen as more just when the task was complex (e.g. performance evaluation). Lee [46] observed that people found human decision-making more fair than algorithmic decision-making when the decision-making related to tasks that were considered to require 'human skills' (e.g. subjective judgement or emotional awareness). Algorithmic and human decision-making were seen as equally fair when the decision involved a task that required 'mechanical skills' such as assignment of work and rostering [46]. Similarly, Castelo et al. [13] found that people viewed algorithms as less effective at subjective tasks compared to objective tasks. In addition to these factors, the way a decision-maker, algorithm or human, is framed in a scenario can impact perceptions [34].

The impact of decisions can also influence preferences for human or algorithmic decision-making. For example, Araujo et al. [6] found that people perceived algorithmic decision-making as fairer than human decision-making in high-stakes decisions made in the health and justice sector contexts. Castelo et al. [13] found that people trusted algorithms less when they were used for tasks that were rated as more consequential. Longoni et al. [54] found that while people generally preferred a human provider for medical decisions, their reluctance to use an algorithmic provider increased when the consequences were high.

The impact of *who* reviews algorithmic decisions on people's perceptions of the *review process* has not yet been explored. The focus of prior research has been on initial decision-making rather than review. Human review of algorithmic decisions is widely advocated for, but as noted by Almada [4], a "trusted third party algorithm" could review an algorithmic decision. There are several reasons why a person may prefer a human reviewer. First, concerns around dignity and individual rights such as autonomy are relevant [4, 32, 60]. Human intervention in the review process is a way to protect rights that people could see as imperiled by algorithmic decision-making [4]. Human review may also be seen as necessary for checking that the algorithm is working as it should [4, 11]. However, human decision-makers, including reviewers, are not perfect. Even judges, the "most esteemed reasoners", can be biased in their decision-making [88]. Depending on its design, algorithmic review might introduce less bias than human review, and will almost certainly be faster, less expensive, and scale better [4, 47]. Even the seemingly simple task of ensuring that the input data is correct may be impossible for a human if the algorithm uses an extensive set of data [12].

We contribute to the literature on perceptions of algorithmic decision-making by extending its scope to study perceptions of algorithmic reviewers in the context of contesting algorithmic decisions.

# 3 STUDY DESIGN AND METHODOLOGY

We explored perceptions of review processes for algorithmic decisions using two different methods that were presented to participants in one Qualtrics survey. Part 1 of the survey focused on eliciting participants' review process *preferences* using a choice-based conjoint analysis. Part 2 of the survey explored perceptions of *fairness* of review processes for algorithmic decisions. Each part of the survey was analysed separately.

## 3.1 Participants

We used the crowdsourcing platform Amazon Mechanical Turk (https://www.mturk.com) to recruit participants. We set the following criteria to qualify for the study: resident of the United States, over 18 years old, high proficiency in English (self rated), and completion of more than 1000 human intelligence tasks with an approval rate above 95% (a commonly used pre-qualification criteria in Mechanical Turk studies [68]). The study was expected to take approximately 25 minutes and participants were paid $5.60 (USD). Ethics approval was provided by the Ethics Committee of our university.

## 3.2 Part 1 - Review process preference using choice-based conjoint analysis

*3.2.1 Choice-based conjoint analysis.* Choice-based conjoint analysis is a method for eliciting preferences [71] and is used in a range of disciplines and settings including marketing [15], litigation [23], and health economics [71]. Conjoint analysis is particularly useful when there are a number of elements that may factor into a person's choice, which makes it a good fit for our study. In a choice-based conjoint analysis, participants are typically asked to choose a preferred option between a number of different *profiles* that each contain *attributes* with different *levels* [15]. This decision will often require the participant to make a trade-off between attributes. For example, when choosing a smartphone, there are numerous attributes that a person might consider including cost, size, colour, brand, storage, camera etc. Each of these attributes will have multiple levels, for example the attribute 'brand', might include Apple, Samsung, and Huawei as levels. Participants generally review a number of different *choice sets* (sets of profiles) during a study. The attributes and levels we use in our study are described in Table 1. Choice-based conjoint analysis reveals which combination of attributes is preferred, how much weight a person places on each attribute when making their decision (*attribute importance*), and the relative importance of the levels within an attribute.

*3.2.2 Attribute selection.* There are many ways that a review process can be designed [55]. Our choice of which elements to focus on was inspired by procedural justice literature and the following review processes for algorithmic decisions suggested by Wachter et al. [84] in relation to Article 22 of the GDPR:

- The review could involve a human making a new decision without the use of any algorithm.
- The review could involve a human making a new decision based on the algorithmic decision and/or the data subject's views on the decision.
- The review could be conducted by a person monitoring the algorithmic decision-making system to ensure it is working correctly, with the algorithmic system providing a new decision.
- A new decision could be made by an algorithmic system without any human intervention.

These review options differ across two key features: who is conducting the review, the 'reviewer'; and the way the review is conducted, the 'type of review'. We incorporated these two key features into our study design.

*Reviewer.* We used two levels for the attribute 'reviewer': *human* and *algorithm*. Work exploring perceptions of algorithmic systems use a variety of terms to describe the systems, including "algorithm" [46], "artificial intelligence" [56], "computer program" [28], and "computer" [45]. In this study we chose to use the terminology 'computer system'. Given that the initial decision in our scenario (outlined below in Section 3.2.3) was made by a computer system, we referred to the computer system reviewer as "Computer System B" to distinguish it from the initial algorithmic decision-maker, "Computer System A".

*Type of review.* The different types of review proposed by Wachter et al. [84] involve the making of a completely new decision (by a human or an algorithm), a new decision that takes into account the previous decision and/or allows the person impacted to participate in the process (the provision of 'voice' as it is referred to in procedural justice literature), and a more technical process involving a review of the initial decision-making system. We adapted these types of review to fit our scenario (see Table 1). The wording of the reviews was pilot tested and refined in a small-scale informal design phase.

*How long the review will take.* In practice, review procedures take varying lengths of time depending on factors such as the complexity of the matter being appealed, resourcing, and the design of the process. We anticipated that the time a review takes will impact the review process participants prefer. So, we introduced a third attribute to analyse: how long the review will take. In comparison to elements of a decision-making process such as 'voice', the timeliness of decision making has received less research attention in the procedural justice literature. Recent research suggests that people prefer processes that are 'timely', but that people feel uncertain about very fast and very slow decision-making [82]. Given the limited research into the impact of time, we used three time periods (7 days, 30 days, and 60 days) to provide a range that differs enough to provide preliminary insight into the impact of this attribute. Our choice of attributes and levels (described in Table 1) resulted in a 2 x 3 x 3 conjoint design.

*3.2.3 Scenario.* We used a hypothetical scenario about a credit loan application that is assessed by an algorithm to introduce participants to the review process options. We chose this context because algorithms are increasingly being used in loan decisions [73]. There are also no established or consistent types of review process in place when people are denied loans, so when people are faced with review options they are less likely to draw on processes they are already familiar with. Participants were presented with the following scenario:

> *Imagine that you have decided to apply for a home loan online. You have carefully filled out the online application form and have uploaded all of the required documents. The home loan provider uses a computer system, called Computer System A, to determine whether you should be provided with a loan. A human is not a part of the decision-making process. Computer System A reviews the information you have provided and predicts how likely you are to repay the loan based on data that the system has collected about thousands of other people who have been provided with home loans. Your application for a home loan has been rejected by Computer System A ('the initial decision'). You want to have this decision reviewed.*

Scenario-based methods are often used in social psychology, justice research, and ethics research to understand perspectives on various issues [46]. Using a scenario allowed us to manipulate the independent variables, which increases the internal validity of the research [43]. The use of scenarios can be criticised for lacking realism. To offset this potential weakness we used a scenario that people generally have some knowledge of and familiarity with, even if they had not applied for a home loan themselves. We also used a first person perspective to increase participant's involvement in the scenario.

## 3.3 Part 2: Fairness perceptions of review processes

*3.3.1 Independent variables.* In Part 2 of the survey we explored participants' perceptions of fairness of different review processes using a within-subjects 2 x 3 design based on two of the variables set out in Table 1: *Reviewer* and *What the Reviewer will do.*

*3.3.2 Dependent variables.*

*Fairness perceptions of review processes.* Noting that there is no universal instrument used to measure perceptions of procedural fairness [74], we used a single question in line with previous research (e.g.[47, 56]); "*Please use the slider to indicate how fair you think the following review processes are (0 means 'not fair at all' and 100 means 'completely fair').*"

*Fairness perceptions of reviewers.* To explore how personal characteristics (such as age, gender, and attitudes towards algorithmic decision-making) impact fairness perceptions of reviewers of algorithmic decision we averaged participants' *fairness perceptions of review process* ratings across review type to calculate average fairness ratings for a human reviewer and for an algorithmic reviewer.

*3.3.3 Subject variables.* Along with gender and age, participants answered the following questions about their attitudes and beliefs about different types of decision-making (the latter four questions are based on [42]):

- how much of an impact the decision to provide someone with a home loan has on that person's life (on a scale of 0-100, 0 = no impact and 100 = high impact);
- who (computer or human) would make the best decision about whether or not to grant a home loan (on a scale of 0-100, 0 = definitely a computer system, 100 = definitely a human);
- how much they have heard about or had experience with a human making loan decisions (on a scale of 0-100, 0 = not at all, 100 = a great deal);
- how much they have heard about or had experience with a computer system making loan decisions (on a scale of 0-100, 0 = not at all, 100 = a great deal); and
- whether they thought computers should make decisions that impact people's lives (on a scale of 0-100, 0 = definitely not, 100 = definitely).

*3.3.4 Quantitative analysis.* We used a factorial ANOVA to analyse fairness perceptions of the different review types and a hierarchical

**Table 1: The attributes and levels used in the choice-based conjoint analysis**

| Attribute | Levels |
|---|---|
| Reviewer | Human<br>Computer System B |
| What the Reviewer will do | The Reviewer will check Computer System A to ensure that the correct information has been taken into account. If it has, the initial decision will be kept. If changes are needed, these will be made and Computer System A will produce a new decision (Verify condition).<br>The Reviewer will make a completely new decision that will be based on the information you submitted with your initial application. No new information will be considered, and the initial decision made by Computer System A will not be taken into account (New Decision condition).<br>The Reviewer will make a completely new decision by taking into account the initial decision made by Computer System A, any objections you make about the initial decision, your original online application and any new information you provide to support your application (New Information condition). |
| How long the review will take | 7 days<br>30 days<br>60 days |

regression analysis to explore the influence of personal characteristics and attitudes on perceptions of fairness of the two different reviewers.

*3.3.5 Qualitative analysis.* Having rated the fairness of the six different review types, we asked participants to use 2-3 sentences to explain why they rated a particular process as the most fair. Participants' responses to this free-text question formed the qualitative data for the study. We analysed the qualitative data from our 100 participants using Braun and Clarke's [10] six-stage approach to reflexive thematic analysis. Taking a contextualist approach to the data, the first author inductively coded submissions using NVivo 12. Based on the coding, initial themes were generated and iteratively refined through a reviewing and writing process.

## 3.4 Procedure

Both parts of the study were presented to participants in one survey, which was developed using Qualtrics. Upon qualifying for and choosing to partake in our study, participants were provided with instructions, a downloadable plain language statement, and a consent form. They were then asked for demographic information (age, gender, and ethnicity).

Part 1 of the survey contained the choice-based conjoint analysis. Participants read the scenario described in Section 3.2.3 and were asked: "Which review option would you choose?". To ensure that participants were not overwhelmed, we presented two review process options, *profiles*, per choice set [24]. An example choice set is shown in Figure 1. We defined 'Computer System B' in the instructions prior to the task as a different computer system to Computer System A (the initial decision-making system). We presented participants with 12 different choice sets. The more choice sets shown to participants the higher the statistical reliability, however this needs to be balanced against respondent fatigue [24]. We used the Qualtrics conjoint analysis software tool, 'Product Optimization (Conjoint)'[2], which uses a randomised balanced design to determine which choice sets were seen by each participant. After completing the 12 choice sets, participants answered two attention check questions. The first asked participants to choose which of two options was a reviewer in the review options they were provided with. Participants who chose the incorrect answer were considered to have failed this attention check, and their data was excluded from analysis. The second attention check asked participants to briefly describe one of the review options they were given. Participants who provided a one-word answer, an answer that did not relate to the question, or an answer that was gibberish were considered to have failed this attention check, and their data was excluded from analysis.

| | **Review option 1** | **Review option 2** |
|---|---|---|
| Reviewer | Computer System B | Human |
| What the Reviewer will do | New decision based on original information: The Reviewer (defined above) will make a completely new decision that will be based on the information you submitted with your initial application. No new information will be considered, and the initial decision made by Computer System A will not be taken into account. | New decision based on additional information: The Reviewer (defined above) will make a completely new decision by taking into account the initial decision made by Computer System A, any objections you make about the initial decision, your original online application and any new information you provide to support your application. |
| How long the review will take | 7 days | 60 days |
| | | |

**Figure 1: An example choice set**

In Part 2 of the survey, participants were first asked to rate the impact of a loan decision on a person's life. They then rated the

[2]https://www.qualtrics.com/au/core-xm/conjoint-analysis/

fairness of the six different base review processes. To address order effects we randomised the order of the review processes. We asked participants to consider the process they rated as most fair and to describe why this was the fairest process. We used this question as another attention check: participants who provided a one-word answer, an answer that did not relate to the question, or an answer that was gibberish were considered to have failed this attention check, and their data was excluded from analysis. Participants then answered questions about who would make the best decision about a loan application, their knowledge of humans and computers making loan decisions, and whether they think computers should make decisions that impact people's lives.

# 4 RESULTS

## 4.1 Participant information

We excluded data from nine participants who failed to appropriately answer the attention check questions (that is they provided an incorrect answer to the first attention check question and/or they provided answers to the free-text questions that were made up of one-word, did not relate to the question or, were gibberish). Of the remaining 100 participants, 62% identified as men and 38% identified as women, with none of the participants choosing non-binary or to self-describe their gender. Participants' ages ranged from 19 to 68 years, with an average of 36.6 years. The average time to complete the study was 17.4 minutes (SD = 9.0).

## 4.2 Part 1: Choice-based conjoint analysis

To answer the research question *'What type of review process do people prefer when contesting an algorithmic decision?'* we conducted a choice-based conjoint analysis using the `mlogit` package [20, 21] in RStudio, which estimates the multinomial logit model underlying the entire sample of 1200 observations (12 binary choices for 100 participants). The `mlogit` package is based on McFadden's work [57, 58] on random utility models [21]. Croissant [21] explains: "These models rely on the hypothesis that the decision maker is able to rank the different alternatives by an order of preference represented by a utility function, the chosen alternative being the one which is associated with the highest level of utility." The result is a multinomial logit model for the whole sample that provides coefficients, which are also referred to as 'preference weights' or 'partworth utilities'.

We dummy coded the variables, which means that each coefficient in Table 2 measures the strength of preference of that level relative to the level of the attribute that was omitted (the reference level, which takes the value of 0). The reference levels in Table 2 are computer system, 30 days, and the New Decision condition. All preference weights were significantly different. Larger estimates indicate stronger preferences and higher utilities indicate more favorable attitudes for that level. Figure 2 visually displays the coefficients/partworth utilities.

These results show that participants preferred a human reviewer to a computer system reviewer, shorter review time periods to longer time periods, the New Information condition, followed by the New Decision condition, then the Verify condition. The preferred review process overall (the highest partworth utility for each level) was the process with a human reviewer, taking 7 days, that uses the New Information condition.

**Table 2: Results of the multinomial logit model**

| Level | Coefficient | SE | p-value |
|---|---|---|---|
| Human reviewer | 1.225 | 0.106 | <0.001 |
| 60 days | -0.944 | 0.121 | <0.001 |
| 7 days | 0.801 | 0.124 | <0.001 |
| Verify condition | -0.551 | 0.125 | <0.001 |
| New Information condition | 1.220 | 0.132 | <0.001 |
| Log-Likelihood: -569.2 | | | |
| McFadden $R^2$: 0.316 | | | |
| $\chi^2$: 524.82 | | | |

The likelihood ratio chi-square test shows that the model is significant ($X^2 = 524.82, p < .001$), which indicates that the inclusion of attribute-level variables in the model significantly improves the model fit in comparison to a model that does not include these variables [31]. The measure of relative model fit, McFadden's pseudo $R^2$, is 0.32: a measure from 0.2 to 0.4 represents a good model fit [31, 59].

*Attribute importance* measures the influence an attribute has in the participant's choice of review process. The relative importance of the three attributes can be gauged from the partworth utility ranges. The greater the importance score, the more weight an attribute carries in the decision-making process. Type of review was the most important attribute to a person when making their choice of review process (37.4%), followed by the time the review takes (36.8%), and then the relative importance of the reviewer (25.8%).

## 4.3 Part 2: Fairness perceptions

*4.3.1 Quantitative analysis.* To explore how different review processes impact people's perceptions of fairness, participants were asked to rate the fairness of the six review processes, which varied by reviewer and type of review. Figure 3 displays the mean fairness rating and standard deviation for each of the six review processes. We used Aligned Rank Transform (ART) [86] to transform the nonparametric factorial data. Common nonparametric tests, such as Friedman, are not able to examine interaction effects when there are multiple factors involved [86]. ART preprocesses data to align it before applying averaged ranks [86]. Following the use of ART to transform our data, we used a factorial ANOVA and then pairwise comparisons to analyse the impact of different review processes on participants' fairness ratings [70].

We found that the type of reviewer has a significant effect on fairness perceptions ($F_{1,495} = 54.27, p < .001$), with human reviewer being rated as more fair than a computer system. The type of review also has a significant effect on perceptions of fairness ($F_{2,495} = 56.99, p < .001$). We conducted pairwise comparisons using the `lsmeans` procedure, which provides p-values that have been corrected for multiple comparisons using Tukey's method [70]. The post hoc tests showed a significant difference between participants' fairness ratings, with higher ratings for the New Information condition compared to the Verify condition ($t_{495} = -7.98, p < .001$),

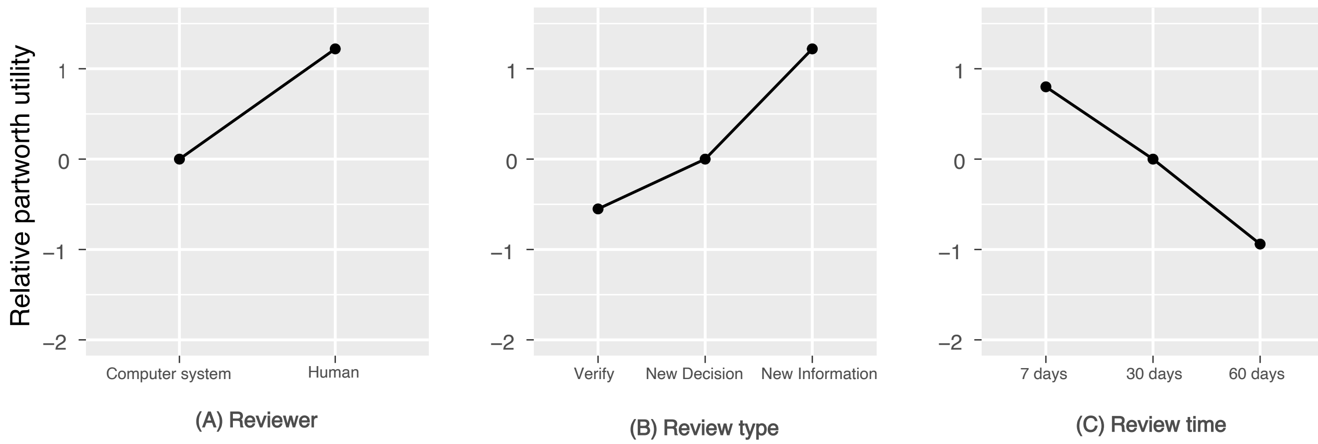


**Figure 2: Conjoint partworth utilities**

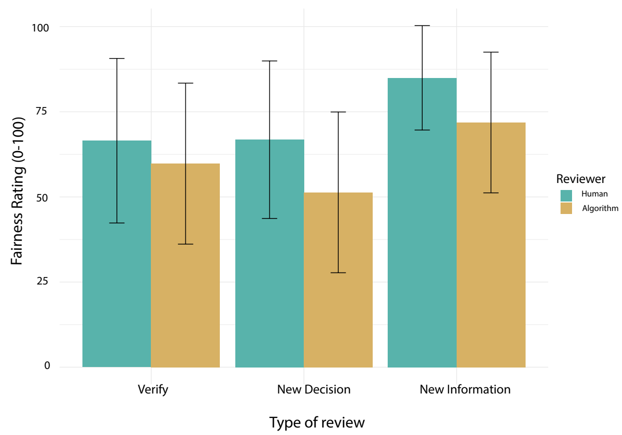


**Figure 3: Fairness ratings of the six base review processes (error bars are standard deviations)**

and higher fairness ratings for the New Information condition compared to the New Decision condition ($t_{495} = -10.13, p < .001$). The difference between fairness ratings of the Verify condition and the New Decision condition was not significant ($p$ = .08). The interaction between reviewer and type of review was not statistically significant ($p$ = .11).

To explore how personal characteristics (such as age, gender, and attitudes towards algorithmic decision-making) impact perceptions of reviewers of algorithmic decisions we averaged participants' ratings of fairness across review types to calculate average fairness ratings for a human reviewer and for an algorithmic reviewer. We explored the influence of personal characteristics and attitudes on these perceptions of fairness using hierarchical regression analysis in two separate models. Table 3 displays the final step (Step 6) of the hierarchical regression analysis for Model 1 (fairness of a human reviewer) and Model 2 (fairness of an algorithmic reviewer).

For Model 1, Step 1 (age and gender) and Step 2 (impact of loan decision) did not produce statistically significant changes in the regression equation. Step 3, adding attitudes about whether a computer or human would make the best loan decision, produced a significant change in the regression model (Δ Adj. $R^2$ = .05, $p$ = 0.01). Step 4 (computers should make important decisions) and Step 5 (experience with humans and experience with computers making loan decisions) did not produce statistically significant changes in the regression equation. Step 6 (fairness rating of an algorithmic reviewer) produced a significant change (Δ Adjusted $R^2$ = .18, $p$ < 0.001). The full regression equation at Step 6 was $R^2$ = .26, Adj. $R^2$ = .19, F(8,91) = 3.98, p < .001.

For Model 2, Step 1 (age and gender), Step 2 (impact of loan decision), and Step 3 (who would make the best loan decision) did not produce statistically significant changes in the regression equation. Adding participant's beliefs that computers should make important decisions in Step 4 produced a significant change (Δ Adj. $R^2$ = .11, $p$ < 0.001). Step 5 did not produce a statistically significant change in the regression equation. Step 6 (fairness rating of a human reviewer) produced a significant change (Δ Adjusted $R^2$ = .17, $p$ < 0.001). The full regression equation at Step 6 was $R^2$ = .32, Adj. $R^2$ = .26, F(8,91) = 5.28, p < .001.

*4.3.2 Qualitative results.* After rating the fairness of each of the six review processes, participants were asked to consider the process they rated as most fair and to describe why this was the fairest process. Responses to this question formed the qualitative data for the study. The four key themes from the thematic analysis are described below.

*Best chance of receiving a different (more favourable) decision.* Many participants viewed the fairest review process as the process that gives the applicant the best chance at receiving **a favourable outcome**, which in this case was an approval for a home loan. To this end, a human was generally preferred as the reviewer rather than a computer system for the following reasons. First, participants thought that a human reviewer was more likely than a computer to make a **different decision**, with the assumption being made that a computer system reviewer would return the same result as the system that made the initial decision (which was a rejection) because of its programming.

**Table 3: Final step of hierarchical regression analyses**

| | Model 1: Human reviewer | | | | Model 2: Algorithmic reviewer | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | *B* | SE | *t* | *p* | *B* | SE | *t* | *p* |
| Age | 0.05 | 0.12 | 0.40 | 0.69 | -0.07 | 0.14 | -0.51 | 0.61 |
| Gender (woman) | -2.67 | 2.74 | -0.97 | 0.33 | 6.62 | 3.35 | 1.98 | 0.05 |
| Impact of loan decision | 0.06 | 0.08 | 0.74 | 0.46 | -0.10 | 0.10 | -1.01 | 0.32 |
| Who would make the best decision about a home loan | 0.15 | 0.07 | 2.20 | **0.03** | -0.01 | 0.09 | -0.15 | 0.88 |
| Whether computers should make important decisions | -0.04 | 0.06 | -0.61 | 0.54 | 0.23 | 0.07 | 3.24 | **0.002** |
| Experience with humans making loan decisions | 0.01 | 0.05 | 0.17 | 0.87 | -0.02 | 0.06 | -0.34 | 0.74 |
| Experience with computers making loan decisions | -0.06 | 0.05 | -1.26 | 0.21 | 0.04 | 0.06 | 0.64 | 0.52 |
| Fairness rating for other reviewer | 0.36 | 0.08 | 4.69 | **<0.001** | 0.55 | 0.12 | 4.69 | **<0.001** |

> *"I felt when we originally had the computer system judge the application (and apparently rejected it), we'd best next have a human submit the application as nearly starting over with the process to see if a different outcome would be achieved...I just don't see how having another computer look over the original outcome would result in a different outcome overall since it will use the same information and come up with the same binary decision."*(P98)

A number of participants characterised the original decision to reject the application as a mistake, although there was no indication in the wording of the question that this was the case. Human reviewers were seen as **less likely to make the same "error"** as the initial computer system, and therefore, more likely to produce a different, more favourable outcome.

> *"One reason I think this is the most fair review process out of the options provided is that a human reviewer is probably less likely to make the same mistake as Computer System A than a different computer system."*(P02)

Many participants saw computers as being **limited by their programming**; unable to take an applicant's individual, unique circumstances into account. In contrast, humans were lauded for their ability to **understand context**, which was seen as providing an applicant with a better chance at receiving a home loan.

> *"I chose this because I believe a human will more accurately be able to interpret my complaints about the initial rejection, understanding the context within which they were made better. Probably more capable of grasping any unique extenuating circumstances that may have lead to it. A totally new review process taking all this into account makes sure I have the best possible chance of success, imho [in my humble opinion]."*(P13)

> *"Computers can only make decisions on whatever they're programmed to do whereas a human can put more critical thought behind the information."*(P31)

Human reviewers were also seen as **more malleable** in their decision-making than computer systems. For instance, Participant 91 stated: *"[H]umans are not so set in stone so they would be able to see where a would be borrower may be able to fit the loan, even if they do not fit a rigid criteria."* The same participant stated that a human reviewer was more likely to see the "potential" of the borrower. In a similar vein, Participant 74 suggested that *"A human can also better tease out which information might be more favorable to the applicant"*.

Others favoured human reviewers due to their **empathy** and their ability to understand the applicant, which was seen as giving the applicant a better chance at approval. For example, Participant 50 suggested that because a human reviewer has emotion they *"will not reject that easily an application."*

The **ability to interact** with a human reviewer was also viewed positively, particularly for those who felt that they had the **ability to influence or persuade the decision-maker**.

> *"I feel I have a better chance of receiving the loan from a human and might be able to grease the wheels by being charming to try to get a loan. I. [sic] cant [sic] do that with a computer system."*(P41)

In addition to the reviewer, how the review was conducted was influential. Some participants felt that there was a better chance of having the loan application approved if a human reviewer was blind to the initial decision: there were concerns that a human would be influenced by the initial rejection and therefore less likely to approve the application. Participant 47 stated: *"I think having the human reviewer examine the systems decision would be too influencing and reduce chances of an approval."*

Many participants saw the **ability to provide additional information** in the review process as important because of its potential to **influence the decision**.

> *"There may be certain circumstances that may arise within someone's credit that could change the mind of the reviewer. The applicant should have the opportunity to adjust or explain all circumstances. The applicant should be able to provide new information and have it considered."*(P89)

> *"New information may provide a different perspective that may help the loan be approved."*(P48)

*The whole story: more than just a number.* A fair appeals process ensures that all of the information relevant to the decision has been taken into account so that the decision is based on a complete understanding of the decision subject and their specific situation. Participant 19 sums up the sentiment of this theme: *"[t]he whole story is not always told by numbers alone."*

A number of participants noted that people have unique situations that should be considered during the decision-making. Many participants saw the ability of humans to understand and **take into account the specific circumstances** of loan applicants and in particular "edge cases" (P74) as essential for making a good decision.

> *"I think that it is most fair when a human reviews my unique situation and all of the relevant information."*(P64)

In contrast to human decision-makers, computer systems were seen as limited by their reliance on "rigid" (P91) programming and their inability take into account the nuances of individual circumstances, particularly if the situation being analysed was an exception to the rules programmed into the system.

> *"Having a human involved for the exceptions is the most fair thing to do. A coder cannot always write an algorithm to go with every life scenario event that may be provided on an application. Computers are great to take care of the things that fall into strict guidelines, but people are still needed to work through the exceptions."*(P63)

In line with the findings of Binns et al. [9], participants wanted to be considered as **whole individuals and not as data points**. Computer systems were seen as number crunchers, whereas human reviewers can take into account information about an applicant and their application beyond the data provided.

> *"A computer program can only really do what it is programed (sic) to do so it cannot be "fair" it's just a set of instructions but a human can make a moral judgement and understands the human aspect of the situation better than a computer ever could."*(P60)

Through the possession of traits such as empathy and intuition, human reviewers were seen as being able to garner important information about the applicant beyond the information a computer system can process.

> *"They [human reviewers] can also see character and determine if this person has a hard work ethic. Computers do not have such an ability."*

Empathy and emotion provided another layer of information to inform the decision.

> *"The human aspect also allows for some judgement calls that I feel that a computer cannot make because it can't feel empathy or connect with the person seeking the loan."*(P44)

While very few participants raised concerns about humans as decision-makers, participant 55 highlighted the potential disadvantage of having a human reviewer, who can be impacted by their own circumstances and emotions: *"Humans can be fallible. They might be having a bad week. They might be fighting with their spouse. It can change things."*

In terms of review process, the majority of participants preferred a process that allowed the applicant to provide all relevant information, enabling the reviewer to make an **informed decision based on a full understanding of the applicant**.

> *"I chose the process run by a human with the most thorough procedure as most fair because this ensures that every piece of information is taken into account and recent developments can be addressed."*(P49)

There was a strong desire for the information being considered to be correct to ensure that the most informed decision was being made. Many participants highlighted the need for new or additional information to be considered, in part because this ensured that the most current information available was being used. Further to this, if any mistakes or errors were made, these could be highlighted and **corrected** through the provision of additional information and through objections made by the applicant.

> *"I think that taking into consideration any new information that may not have been available during the initial decision is fair because the original decision was not making its decision on current facts."*(P22)

*Human intervention: the best of both worlds.* Most participants accepted the use of a computer system to make the initial decision about the loan, but made it clear that there should be **human involvement at some stage in the process**. For some, a human review offered a different but complementary point of view to the algorithmic decision; a second set of eyes. A number of participants stated that having both a computer system and human involved in the decision-making process would result in a better decision, especially given that neither decision-maker is perfect.

> *"I feel it is the fairest because I would get two different reviews by two different systems - one via a human brain and the other via a computer. If both a human and a computer came to the same conclusion about the loan, even after getting more information, then it is likely that the decision was correct or at least justified."*(P87)

Many participants liked this idea of having "two means of verification" (P11). A number of participants characterised a human review as a way to double-check the decision, or as providing a second chance or second opinion.

> *"I think that after having a computer system make the initial decision I think its only fair to have a human double check to make sure and the information is correct and that if any new information needing to be put in."*(P36)

A frequent concern with the sole use of a computer system as decision-maker and reviewer was that any errors would be systematic and therefore repeated. Participants were confident in a human reviewer's ability to see any mistakes made by the computer system, including whether the right information was taken into account, whether the system was functioning as designed, and ultimately whether the right decision had been reached.

> *"I think having some human intervention is good. I believe that letting an algorithm do all the work without intervention could lead to uncaught errors that may cause a loan to be rejected. Computers can be wrong and humans can be wrong. If we work together we can produce better results."*(P69)

Some participants preferred to have a human reviewer make a completely fresh decision without influence from the initial decision.

This style of review effectively amounts to the decision being made by a human, with no computer system in the loop.

> *"I think having the human reviewer examine the systems decision would be too influencing and reduce chances of an approval."*(P47)

A number of participants saw an appeal process involving a human reviewer as fairest because of their high regard for humans as decision-makers.

> *"I think any system using a human is inherently more fair than one that uses a computer because human thinking and intuition is still superior in this regard."*(P61)

*Preserving dignity.* For many participants the fairest process was the one that treated loan applicants with **dignity and respect as individuals**. For example, participants tended to favour human reviewers over computers because they did not want to be seen as "just numbers" (P19, P62, P91); they want a 'humane' (P07) appeal process with a "genuine touch" (P19). As outlined above in the theme 'The whole story: more than just a number' there was also a strong desire that each individual's situation be understood and taken into account in the decision-making.

> *"I think that having a human reviewing the computer work would make us feel like we weren't just numbers, but that they considered our specific situation."*(P62)

A couple of participants highlighted the indignity and frustration that a person would feel if they contested a decision made by a computer system only to have their appeal directed to another computer system.

> *"If I was already rejected by a computer system, I do not want another computer system to make a decision."*(P05)

The impact of the decision being made, to provide a loan or not, was a factor that some participants took into account in determining the fairest review process. A loan decision was seen as having **a real impact on a person's life**, with a number of participants suggesting that a decision with this kind of impact required human review.

> *"Home loans can have huge impacts on people life so it's good to have a human element to it, someone who shows compassion."*(P90)

The ability to interact with someone who can provide an explanation or advice was also seen as an important element of the review process.

> *"With something like a home loan, sometimes it's important to be able to appeal to someone who can reason with you and be level with you about what went wrong or what you need to actually DO in order to get a positive result."*(P88)

Indeed, many participants wanted the **ability to participate** in the decision-making process, emphasising the importance of the loan applicant taking an active role in a decision that impacts them. This notion, of participation and the ability to express opinions and provide evidence, is referred to as 'voice' in the procedural justice literature.

> *"They also take into account objections that the person impacted by the decision has and new information provided. So the person impacted will have some say in the matter and another chance if the loan is rejected."*(P73)

A number of participants commented that their objections would be understood better by a human in comparison to a computer. This is in line with procedural justice literature on voice that suggests that people need to be able to **express their views but also feel that they have been heard** [49].

> *"I think the human making a new decision based on original info plus new info would be the most fair. The human would have a better ability to understand why various objections were relevant, especially if they were uncommon or unexpected (a computer can only take into account what it's programmers thought of ahead of time, and can't really adapt to unexpected information)."*(P53)

Overall, what the qualitative data shows is that participants' impressions of different review processes are strongly influenced by their ability to get the decision that they want, and less around getting a good or fair decision. That is, participants primarily discuss ideas around influencing the decision to get the loan rather than whether they could repay the loan. In this sense, the participants' view of the process is one of self interest, not taking into account the fairness to the organisation granting the loan or the other applicants who would miss out if the participant's loan was granted.

# 5 DISCUSSION

When it comes to contesting algorithmic decisions, the preferred review process for our participants is a timely review conducted by a human, taking into account the initial information provided, the initial decision, new information, and any objections. Human reviewers were also seen as more *fair* than computer system reviewers, and the New Information condition was rated as the *fairest* type of review. These two analyses indicate that the type of review process participants prefer aligns with the review process they perceive as most fair. In this section we consider the design implications of our findings.

## 5.1 The ability to participate and feel heard

While human reviewers were strongly preferred when the type of review and review time were held constant, we found that when forced into a trade-off situation in the conjoint analysis, participants placed more weight on the type of review and the review time than the reviewer. As an example, when faced with two different review processes, both taking 7 days, but one with a human reviewer and the Verify condition, and one with a computer system and the New Information condition, participants were more likely to choose the process with the New Information style of review and the algorithmic reviewer. So, while guidelines considering algorithmic decision-making often set out the need for human review after algorithmic decision-making (e.g. [26, 66]), our analysis suggests that decision subjects are willing to sacrifice human review for a more timely review or a process that allows for more participation and voice. Thus, when designing a review process each of the three attributes we have explored deserve careful consideration.

Of the three different types of review, participants preferred the type of review that allowed them the most leeway to participate–to have their say, and put forward their best case. This finding aligns with previous research on 'process control', also referred to as 'voice', in the procedural justice literature [77]. There are many ways to allow for voice during a decision-making process. In this study, the review process allowed for extra information to be submitted and for objections to be made. The ease at which these actions can be taken will impact the contestation experience. For example, some decision subjects may prefer verbal over written objections. If objections need to be written, whether they can be submitted via an online interface, over email, or via a letter will also impact a person's experience of the review process. How long these steps take must also be considered given the importance participants placed on receiving a timely review.

Procedural justice literature sets out that in addition to being able to exercise voice, decision subjects must also feel as though they have been 'heard' [49]. Based on our qualitative results, participants view algorithms as far less capable than humans of parsing objections or understanding the nuance of people's individual situations. As such, ensuring that a decision subject feels heard may be a hurdle that algorithmic reviewers struggle to meet. In this study, we focused on asynchronous review processes, where an appeal occurs after a decision has been made. Work on designing for contestability envisions algorithmic decision-making systems that support interactivity, where a person impacted by a decision can interact in-band (e.g. [33, 41, 61]). For example, an algorithmic decision-making system could be designed to allow a person subject to decision-making to access information about decision inputs or conclusions via an interface that allows them to lodge disagreement with system, like the psychotherapy tool designed by [33]. This kind of interaction provides decision subjects with knowledge about how the system works, the information it takes into account, and the ability to exercise some control over the decision. A system could also provide decision subjects with the ability to correct or update information to ensure that decisions are based on accurate information. Being able to interact with a system and see it react and adjust to new information or objections may satisfy the desire to feel heard.

## 5.2 Algorithmic aversion and the potential for explanation

Our results indicate that all else being equal, humans are preferred over algorithms as reviewers of algorithmic loan decisions and they are also seen as more fair. These results lend support to the notion of algorithmic aversion. However, it is important to highlight that we explore perceptions of algorithmic and human reviewers in the context of appealing an algorithmic decision, which is a different context to work investigating perceptions of algorithmic decision-making, which focuses on decision-making in the first instance. Importantly, our scenario involved an algorithm making an initial decision to reject a loan application. Against this backdrop, many participants viewed a human reviewer as more likely to provide a *different, more favourable* decision compared to an algorithm, and thus favoured a human reviewer. Interestingly, our qualitative results suggest that participants were generally tolerant of receiving an initial algorithmic decision, but the majority were keen to have a human in the decision-making process as reviewer.

Additionally, in line with previous studies where the perceived capabilities of, and attitudes towards, the decision-maker contribute to algorithmic aversion (e.g. [13, 38, 42, 54]), we found that the belief that humans would make better loan decisions than a computer is associated with higher perceived fairness of human reviewers and the belief that computers should make important decisions that impact people's lives is associated with higher perceptions of fairness of a computer system reviewer. These results from the hierarchical regression are also reflected in the findings from our qualitative analysis. For example, a number of participants saw decision-making by computer systems as rigid, reliant on how they have been programmed, and not able to easily adapt to information that does not fit within pre-programmed rules. As such, computer systems were seen as restricted in their ability to consider 'edge cases' or applications with unique situations. This highlights potential misconceptions about the abilities of algorithmic systems: many algorithms now being used in decision-making are not rule based systems that have been programmed by people, but are based on statistical techniques using machine learning to consider vast amounts of data to come to a decision and can handle more attributes about a decision than a person can. These results suggest that if an algorithm is used as a decision-maker or a reviewer, clearly explaining how the system works would be beneficial.

Interestingly, recent work exploring perceptions of algorithmic decision-making has shown that beliefs about capabilities can be altered, which in turn impacts perceptions of algorithmic decision-makers. For example, Longoni et al. [54] found that aversion to medical AI can be decreased by explicitly stating that the AI can provide personalised advice, while Castelo et al. [13] found that algorithmic aversion decreased when a task was rephrased to sound more objective (and therefore, more capable of being performed by an algorithm). These findings highlight the impact of the explanations used to describe algorithmic decision-making. While these findings can be used to inform the way that review processes are described to decision subjects in an effort to decrease algorithmic aversion they also highlight the concerning potential to manipulate people through the use of careful phrasing in explanations.

## 5.3 The relevance of procedural justice in designing algorithmic appeal processes

In line with findings from the procedural justice literature relating to 'outcome favourability' (e.g. [43, 78]), we found that participants judged the review process that they believed would result in them being granted a loan as the most fair. However, the idea that everyone subject to decision-making can achieve a favourable outcome is not realistic. For example, when applying for a job there are a limited number of interviews available, so a large percentage of people who apply for the job will not progress to interview. In the context of a loan, lenders have a finite amount of funds to lend out, as well as the responsibility to ensure that borrowers are able to repay the loan. Indeed, a willingness to provide credit, especially when it should not have been offered, was one of the key factors contributing to the global financial crisis [64]. So, while understanding the perspectives and needs of people who will be impacted by the decision is essential, it is also important to understand and take into account the very human desire for a favourable outcome. Thus, perspectives about what would make a fair review procedure should also be obtained from additional stakeholders including decision-makers and neutral third parties.

In addition to outcome favourability, our results highlight the importance of having a voice in the decision-making process, that decisions are based on accurate information, and that decision subjects are treated with dignity, all of which have been found to contribute to feelings of justice in human decision-making [17, 48]. These synergies suggest that the procedural justice literature, which largely focuses on the fairness of human decision-making, can be drawn from when designing algorithmic decision-making procedures. However, there may be challenges when adapting relevant findings from procedural justice work to algorithmic decision-making. For example, dignity is a complex concept, and whether having an algorithmic system making high consequence decisions will ever be perceived as "dignified" warrants further exploration. Ultimately, findings from procedural justice research suggest that the whole decision-making process needs to be carefully considered and designed, not just the appeal process.

## 6 LIMITATIONS AND FUTURE WORK

The scenario that we used for the choice-based conjoint analysis was hypothetical. Although this is a standard method used to explore perceptions and attitudes in relation to algorithmic decision-making (e.g. [9, 47, 54]), it lacks external validity; the participants were not being faced with a real loan application rejection and so the consequences and significance of the decision are lacking. Despite this limitation, there are a number of studies that indicate that people's responses and behaviour in controlled studies are similar to how they act in real life [87].

The terminology we used in the scenario to describe the algorithmic system (i.e. "computer system") may have impacted people's perceptions of the system. Interestingly, Langer et al. [44] found that while the terminology used to describe algorithmic systems impacts people's fairness perceptions and feelings of trust in relation to the system, the use of different terminology (e.g. "artificial intelligence", "computer program", "algorithm") did not impact perceptions of the ability of the algorithmic system when that system was being compared directly with a human decision-maker, which is the type of comparison we made in our work.

Further to this, the algorithmic decision in our scenario was a rejection of the loan application. This outcome may have primed participants to view algorithms negatively, thus impacting their perceptions of algorithmic reviewers. However, given that the context we are exploring is the appeal of algorithmic decisions, having an initial negative algorithmic decision is realistic as people are unlikely to contest a decision that is favourable to them. Indeed, our qualitative results indicate that one of the many reasons people preferred human review was because the algorithmic decision being appealed was a rejection. This is an important consideration for decision-makers when designing a review process.

One of the review types that we explored was the New Information condition. This style of review took into account original information, new information, objections, and the algorithmic decision. This was the preferred style of review, however it is not clear whether one of its affordances was more important to participants than another. These nuances of 'voice' could be explored in greater detail by isolating them, so that one review process involves the ability to provide new information and another review process provides the ability to object. In a similar vein, the type of reviewer could be specified in more detail. For example, the review may be run by an algorithm developed by a government body, a university research institute, or a private company. Additional attributes of a review process that may impact a person's choice could also be investigated (e.g. cost, longer/shorter time frames).

We used a within-subject design to explore perceptions of fairness of different review processes for algorithmic decisions. This meant that participants were able to compare across various types of review. In reality, people will be offered one type of review and would not be comparing between options. A between-subjects design testing one type of review per condition might yield different results.

We only explored one context (loan decisions) using a mid-sized sample of participants from Mechanical Turk. Our findings may not generalise to different contexts. Future studies could explore whether these results can be replicated using a between-subject design and across algorithmic decisions in different settings, such as hiring decisions or performance evaluation. Further, we did not explore cultural differences. The "right to appeal" is a Western, democratic notion and results may differ across cultures.

## 7 CONCLUSION

With algorithms increasingly making decisions that significantly impact people's lives, the ability to contest such decisions is gaining attention. There are no guidelines around how appeal processes for algorithmic decisions should be designed. In this work we explore perceptions of different types of processes that can be used to appeal algorithmic decisions. We consider perceptions from two angles: first, through a conjoint analysis where we force a choice between review processes; and second, by asking participants to rate the fairness of six different review processes. We find that participants prefer timely review processes with a human reviewer that allow them 'voice' through the provision of new information and the ability to make objections to the initial decision. Our qualitative analysis provides an understanding of what participants look for in a process they consider to be fair: the chance to change the decision to a more favourable outcome, being treated with dignity, human involvement, and having their unique situation taken into account.

## ACKNOWLEDGMENTS

We would like to thank our reviewers for their valuable feedback. This research was partly funded by Australian Research Council Discovery Grant DP190103414. Henrietta Lyons is supported by the Faculty of Engineering and Information Technology Ingenium scholarship program. Eduardo Velloso is the recipient of an Australian Research Council Discovery Early Career Researcher Award (Project Number: DE180100315) funded by the Australian Government.

# REFERENCES


[1] J Stacy Adams. 1965. Inequity in Social Exchange. In *Advances in Experimental Social Psychology*. Vol. 2. Elsevier, 267–299.

[2] Richard Adams and Heather Stewart. 2020. Ofqual 'blindsided' government by revoking A-level appeals process. Retrieved May 4, 2021 from https://www.theguardian.com/education/2020/aug/16/ofqual-blindsided-government-by-revoking-a-level-appeals-process.

[3] Sheldon Alexander and Marian Ruderman. 1987. The role of procedural and distributive justice in organizational behavior. *Social Justice Research* 1, 2 (1987), 177–198.

[4] Marco Almada. 2019. Human Intervention in Automated Decision-making: Toward the Construction of Contestable Systems. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law (ICAIL '19)*. 2–11.

[5] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. Retrieved 24 July 2020 from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[6] Theo Araujo, Natali Helberger, Sanne Kruikemeier, and Claes H De Vreese. 2020. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society* (2020), 1–13.

[7] Judith Bannister, Anna Olijnyk, and Stephen McDonald. 2018. *Government accountability: Australian administrative law* (2nd. ed.). Cambridge University Press, Port Melbourne, Vic.

[8] Emre Bayamlıoğlu. 2021. The right to contest automated decisions under the General Data Protection Regulation: Beyond the so-called "right to explanation". *Regulation & Governance* (2021).

[9] 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. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, 1–14. https://doi.org/10.1145/3173574.3173951

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3 (2006), 77–101.

[11] Kiel Brennan-Marquez and Stephen Henderson. 2019. Artificial Intelligence and Role-Reversible Judgment. *Journal of Criminal Law and Criminology* 109, 1, Article 2 (2019), 27 pages.

[12] Maja Brkan. 2019. Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. *International Journal of Law and Information Technology* 27, 2 (2019), 91–121.

[13] Noah Castelo, Maarten W Bos, and Donald R Lehmann. 2019. Task-dependent algorithm aversion. *Journal of Marketing Research* 56, 5 (2019), 809–825.

[14] Danielle Keats Citron and Frank Pasquale. 2014. The scored society: Due process for automated predictions. *Wash. L. Rev.* 89 (2014), 1.

[15] Ludovik Coba, Laurens Rook, Markus Zanker, and Panagiotis Symeonidis. 2019. Decision making strategies differ in the presence of collaborative explanations: two conjoint studies. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 291–302.

[16] Jason A Colquitt, Donald E Conlon, Michael J Wesson, Christopher OLH Porter, and K Yee Ng. 2001. Justice at the Millennium: A Meta-Analytic Review of 25 Years of Organizational Justice Research. *Journal of Applied Psychology* 86, 3 (2001), 425.

[17] Jason A Colquitt and Jessica B Rodell. 2015. Measuring justice and fairness. In *The Oxford Handbook of Justice in the Workplace*, Russell S. Cropanzano and Maureen L. Ambrose (Eds.). Oxford University Press.

[18] Donald E Conlon. 1993. Some tests of the self-interest and group-value models of procedural justice: Evidence from an organizational appeal procedure. *Academy of Management Journal* 36, 5 (1993), 1109–1124.

[19] Courts and Tribunals Judiciary (United Kingdom). 2020. The right to appeal. Retrieved on 31 March 2020 from https://www.judiciary.uk/about-the-judiciary/the-judiciary-the-government-and-the-constitution/jud-acc-ind/right-2-appeal/.

[20] Yves Croissant. 2012. Estimation of multinomial logit models in R: The mlogit Packages. *R package version 0.2-2. URL: http://cran. r-project. org/web/packages/mlogit/vignettes/mlogit.pdf* (2012).

[21] Yves Croissant. 2020. Estimation of Random Utility Models in R: The mlogit Package. *Journal of Statistical Software* 95, 11 (2020), 1–41.

[22] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114.

[23] Felix Eggers, John R Hauser, and Matthew Selove. 2016. The Effects of Incentive Alignment, Realistic Images, Video Instructions, and Ceteris Paribus Instructions on Willingness to Pay and Price Equilibria. In *Proceedings of the Sawtooth Software Conference*.

[24] Felix Eggers, Henrik Sattler, Thorsten Teichert, and Franziska Völckner. 2018. Choice-Based Conjoint Analysis. In *Handbook of Market Research*. Springer.

[25] Theodoros Evgeniou, David R. Hardoon, and Anton Ovchinnikov. 2020. What Happens When AI is Used to Set Grades? Retrieved May 4, 2021 from https://hbr.org/2020/08/what-happens-when-ai-is-used-to-set-grades.

[26] Sophie Farthing, John Howell, Katerina Lecchi, Zoe Paleologos, Phoebe Saintilan, and Edward Santow. 2021. Human Rights and Technology (Australian Human Rights Commission). https://humanrights.gov.au/our-work/rights-and-freedoms/publications/human-rights-and-technology-final-report-2021.

[27] Jerald Greenberg. 1986. Determinants of perceived fairness of performance evaluations. *Journal of Applied Psychology* 71, 2 (1986), 340.

[28] Nina Grgić-Hlača, Christoph Engel, and Krishna P. Gummadi. 2019. Human Decision Making with Machine Advice: An Experiment on Bailing and Jailing. *Proceedings of ACM Hum.-Comput. Interact.* 3, CSCW, Article 178 (2019), 25 pages.

[29] Nina Grgić-Hlača, Elissa M. Redmiles, Krishna P. Gummadi, and Adrian Weller. 2018. Human Perceptions of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*. 903–912. https://doi.org/10.1145/3178876.3186138

[30] Nina Grgić-Hlača, Muhammad Bilal Zafar, Krishna P. Gummadi, and Adrian Weller. 2018. Beyond Distributive Fairness in Algorithmic Decision Making: Feature Selection for Procedurally Fair Learning. In *Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*. 51–60.

[31] A Brett Hauber, Juan Marcos González, Catharina GM Groothuis-Oudshoorn, Thomas Prior, Deborah A Marshall, Charles Cunningham, Maarten J IJzerman, and John FP Bridges. 2016. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. *Value in Health* 19, 4 (2016), 300–315.

[32] Mireille Hildebrandt. 2019. Privacy as protection of the incomputable self: From agnostic to agonistic machine learning. *Theoretical Inquiries in Law* 20, 1 (2019), 83–121.

[33] Tad Hirsch, Kritzia Merced, Shrikanth Narayanan, Zac E. Imel, and David C. Atkins. 2017. Designing Contestability: Interaction Design, Machine Learning, and Mental Health. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. 95–99.

[34] Yoyo Tsung-Yu Hou and Malte F Jung. 2021. Who is the expert? Reconciling algorithm aversion and algorithm appreciation in AI-supported decision making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–25.

[35] AI Now Institute. 2018. Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems. Retrieved 15 December 2019 from https://ainowinstitute.org/litigatingalgorithms.pdf.

[36] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in ai. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 624–635.

[37] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1 (2019), 389–399.

[38] Ekaterina Jussupow, Izak Benbasat, and Armin Heinzl. 2020. Why are we averse towards algorithms? A comprehensive literature review on Algorithm aversion. In *Twenty-Eighth European Conference on Information Systems (ECIS2020)*.

[39] Margot Kaminski. 2019. Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability. *Southern California Law Review* 92, 6 (2019), 1529–1616.

[40] Margot E Kaminski and Jennifer M Urban. 2021. The Right to Contest AI. *Columbia Law Review* 121, 7 (2021).

[41] Daniel Kluttz and Deirdre K Mulligan. 2020. Automated decision support technologies and the legal profession. *Berkeley Technology Law Journal* 34, 3 (2020).

[42] 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* (New Orleans, LA, USA) *(AIES '18)*. 204–209. https://doi.org/10.1145/3278721.3278752

[43] David A Kravitz, Eugene F Stone-Romero, and Jeffrey A Ryer. 1997. Student evaluations of grade appeal procedures: The importance of procedural justice. *Research in Higher Education* 38, 6 (1997), 699–726.

[44] Markus Langer, Tim Hunsicker, Tina Feldkamp, Cornelius J König, and Nina Grgić-Hlača. 2021. " Look! It's a Computer Program! It's an Algorithm! It's AI!": Does Terminology Affect Human Perceptions and Evaluations of Intelligent Systems? *arXiv preprint arXiv:2108.11486* (2021).

[45] Markus Langer, Cornelius J König, and Victoria Hemsing. 2020. Is anybody listening? The impact of automatically evaluated job interviews on impression management and applicant reactions. *Journal of Managerial Psychology* (2020).

[46] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (2018), 1–16.

[47] Min Kyung Lee, Anuraag Jain, Hea Jin Cha, Shashank Ojha, and Daniel Kusbit. 2019. Procedural Justice in Algorithmic Fairness: Leveraging Transparency and Outcome Control for Fair Algorithmic Mediation. *Proceedings ACM Hum.-Comput. Interact.* 3, CSCW, Article 182 (Nov. 2019), 26 pages. https://doi.org/10.1145/3359284

[48] Gerald Leventhal. 1980. What Should Be Done With Equity Theory? New Approaches to the Study of Fairness in Social Relationships. In *Social Exchange*, Kenneth J Gergen, Martin S Greenberg, and Richard Hartley (Eds.). Plenum Press.

[49] E. Allan Lind and Christiane Arndt. 2016. *Perceived Fairness and Regulatory Policy: A Behavioural Science Perspective on Government-Citizen Interactions.* OECD

Regulatory Policy Working Papers 6. OECD Publishing. https://doi.org/10.1787/1629d397-en

[50] E Allan Lind, Ruth Kanfer, and P Christopher Earley. 1990. Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments. *Journal of Personality and Social Psychology* 59, 5 (1990), 952.

[51] E Allan Lind, Robert J MacCoun, Patricia A Ebener, William LF Felstiner, Deborah R Hensler, Judith Resnik, and Tom R Tyler. 1990. In the eye of the beholder: Tort litigants' evaluations of their experiences in the civil justice system. *Law and Society Review* 24, 4 (1990), 953–996.

[52] E Allan Lind and Tom R Tyler. 1988. *The social psychology of procedural justice.* Springer US.

[53] Jennifer M Logg, Julia A Minson, and Don A Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151 (2019), 90–103.

[54] Chiara Longoni, Andrea Bonezzi, and Carey K Morewedge. 2019. Resistance to medical artificial intelligence. *Journal of Consumer Research* 46, 4 (2019), 629–650.

[55] Henrietta Lyons, Eduardo Velloso, and Tim Miller. 2021. Conceptualising Contestability: Perspectives on Contesting Algorithmic Decisions. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–25.

[56] Frank Marcinkowski, Kimon Kieslich, Christopher Starke, and Marco Lünich. 2020. Implications of AI (Un-)Fairness in Higher Education Admissions: The Effects of Perceived AI (Un-)Fairness on Exit, Voice and Organizational Reputation. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. 122–130. https://doi.org/10.1145/3351095.3372867

[57] Daniel McFadden. 1973. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, P. Zarembka (Ed.). Academic Press: New York, 105–142.

[58] Daniel McFadden. 1974. The measurement of urban travel demand. *Journal of Public Economics* 3, 4 (1974), 303–328.

[59] Daniel McFadden. 1977. Quantitative methods for analyzing travel behavior of individuals: some recent developments. In *Behavioural Travel Modelling*, David A Hensher and Peter R Stopher (Eds.). Vol. 474. Croom Helm London:London, 279–318.

[60] Isak Mendoza and Lee A. Bygrave. 2017. The Right not to be Subject to Automated Decisions based on Profiling. In *EU Internet Law*, Tatiani Synodinou, Philippe Jougleux, Christiana Markou, and Thalia Prastitou (Eds.). Springer.

[61] Deirdre K. Mulligan, Daniel Kluttz, and Nitin Kohli. 2020. Shaping Our Tools: Contestability as a Means to Promote Responsible Algorithmic Decision Making in the Professions. In *After the Digital Tornado: Networks, Algorithms, Humanity*, Kevin Werbach (Ed.). New York: Cambridge University Press.

[62] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 2, 11 (2018), 4366–4383.

[63] Rosanna Nagtegaal. 2021. The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly* 38, 1 (2021), 101536.

[64] Reserve Bank of Australia. Undated. The Global Financial Crisis. Retrieved December 21, 2021 from https://www.rba.gov.au/education/resources/explainers/the-global-financial-crisis.html.

[65] Australian Government Department of Industry Innovation and Science. 2019. AI Ethics Framework. Retrieved on 6 December 2019 from https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles.

[66] Independent High-Level Expert Group on Artificial Intelligence. 2019. Ethics Guidelines for Trustworthy AI. Retrieved 24 February 2020 from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[67] Sonja K Ötting and Günter W Maier. 2018. The importance of procedural justice in human–machine interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior* 89 (2018), 27–39.

[68] Eyal Peer, Joachim Vosgerau, and Alessandro Acquisti. 2014. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods* 46, 4 (2014), 1023–1031.

[69] Rashida Richardson, Jason M. Schultz, and Vincent M. Southerland. 2019. Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems. Retrieved 20 December 2019 from https://ainowinstitute.org/litigatingalgorithms-2019-us.pdf.

[70] Judy Robertson and Maurits Kaptein. 2016. *Modern statistical methods for HCI.* Springer.

[71] Mandy Ryan and Shelley Farrar. 2000. Using conjoint analysis to elicit preferences for health care. *BMJ* 320, 7248 (2000), 1530–1533.

[72] Claudio Sarra. 2020. Put Dialectics into the Machine: Protection against Automatic-decision-making through a Deeper Understanding of Contestability by Design. *Global Jurist* 20, 3 (2020).

[73] Nripsuta Ani Saxena, Karen Huang, Evan DeFilippis, Goran Radanovic, David C. Parkes, and Yang Liu. 2019. How Do Fairness Definitions Fare? Examining Public Attitudes Towards Algorithmic Definitions of Fairness. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Honolulu, HI, USA) *(AIES '19)*. 99–106. https://doi.org/10.1145/3306618.3314248

[74] Ayelet Sela. 2018. Can Computers Be Fair: How Automated and Human-Powered Online Dispute Resolution Affect Procedural Justice in Mediation and Arbitration. *Ohio St. J. on Disp. Resol.* 33 (2018), 91.

[75] Steven Shavell. 1995. The appeals process as a means of error correction. *The Journal of Legal Studies* 24, 2 (1995), 379–426.

[76] Blair H Sheppard. 1985. Justice is no simple matter: Case for elaborating our model of procedural fairness. *Journal of Personality and Social Psychology* 49, 4 (1985), 953.

[77] John Thibaut and Laurens Walker. 1975. *Procedural Justice: A Psychological Analysis.* Lawrence Erlbaum Associates, Hillsdale, NJ.

[78] Tom R Tyler. 1988. What is procedural justice-criteria used by citizens to assess the fairness of legal procedures. *Law & Soc'y Rev.* 22 (1988), 103.

[79] United States District Court 2017. Houston Federation of Teachers, Local 2415, et al v Houston Independent School District, 251 F.Supp.3d 116. leagle.com/decision/infdco20170530802.

[80] Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. "At the End of the Day Facebook Does What It Wants" How Users Experience Contesting Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–22.

[81] Kristen Vaccaro, Ziang Xiao, Kevin Hamilton, and Karrie Karahalios. 2021. Contestability For Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–28.

[82] Annukka Valkeapää and Tuija Seppälä. 2014. Speed of decision-making as a procedural justice principle. *Social Justice Research* 27, 3 (2014), 305–321.

[83] Niels van Berkel, Jorge Goncalves, Danula Hettiachchi, Senuri Wijenayake, Ryan M. Kelly, and Vassilis Kostakos. 2019. Crowdsourcing Perceptions of Fair Predictors for Machine Learning: A Recidivism Case Study. *Proc. ACM Hum.-Comput. Interact.* 3, 28 (2019), 21 pages. https://doi.org/10.1145/3359130

[84] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2018. Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law and Technology* 31, 2 (2018), 841–887.

[85] Ruotong Wang, F. Maxwell Harper, and Haiyi Zhu. 2020. Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences. In *Proceedings of the 2020 Conference on Human Factors in Computing Systems* (Honolulu, HI, USA). 1–14. https://doi.org/10.1145/3313831.3376813

[86] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. 143–146. https://doi.org/10.1145/1978942.1978963

[87] Sarah Woods, Michael Walters, Kheng Lee Koay, and Kerstin Dautenhahn. 2006. Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach. In *9th IEEE International Workshop on Advanced Motion Control, 2006.* IEEE, ..., 750–755.

[88] John Zerilli, Alistair Knott, James Maclaurin, and Colin Gavaghan. 2019. Transparency in algorithmic and human decision-making: Is there a double standard? *Philosophy & Technology* 32, 4 (2019), 661–683.