International Journal of
Human–Computer Interaction

Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Fairness Perceptions of Artificial Intelligence: A Review and Path Forward

Devesh Narayanan, Mahak Nagpal, Jack McGuire, Shane Schweitzer & David De Cremer

To cite this article: Devesh Narayanan, Mahak Nagpal, Jack McGuire, Shane Schweitzer & David De Cremer (2024) Fairness Perceptions of Artificial Intelligence: A Review and Path Forward, International Journal of Human–Computer Interaction, 40:1, 4-23, DOI: 10.1080/10447318.2023.2210890

To link to this article: https://doi.org/10.1080/10447318.2023.2210890

Published online: 26 May 2023.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
2024, VOL. 40, NO. 1, 4–23
https://doi.org/10.1080/10447318.2023.2210890

Taylor & Francis
Taylor & Francis Group

Fairness Perceptions of Artificial Intelligence: A Review and Path Forward

Devesh Narayanan, Mahak Nagpal, Jack McGuire, Shane Schweitzer, and David De Cremer

Centre on AI Technology for Humankind, NUS Business School, National University of Singapore, Singapore

ABSTRACT

A key insight from research on organizational justice is that fairness is in the eye of the beholder. With increasing discussions – especially among computer scientists and policymakers – about the potential biases and unfairness of decisions made by Artificial Intelligence (AI) systems, there is a critical need to consider how decision-subjects perceive the fairness of AI-led decision-making. Drawing upon theoretical and empirical perspectives on perceived fairness in organizational justice scholarship, this review categorizes and analyzes perceptions of AI fairness as they impact the effective implementation of AI in workplaces and beyond. Specifically, we review existing empirical research on AI fairness according to distinct dimensions of perceived fairness – distributive, procedural, interpersonal, and informational – with a focus on its potential to inform organizational decision-making. In doing so, we provide new insights and offer directions for future interdisciplinary research in this burgeoning field.

KEYWORDS

Perceived fairness; artificial intelligence; organizational justice

1. Introduction

Organizations are a key site for the development, deployment, and management of modern Artificial Intelligence (AI) systems. Indeed, AI systems are increasingly being adopted by organizations in critical decision-making contexts: both internally, in managerial processes such as the allocation of payment, rewards, tasks, shifts, promotions, and training (Raisch & Krakowski, 2021; Wilson & Daugherty, 2018), as well as externally, in the delivery of AI-augmented products and services to customers (Fountaine et al., 2019; Benbya et al., 2020). Since such decisions are often sensitive and high-stakes, there have been mounting concerns about the fairness of AI-augmented decision-making (De Cremer, 2020). Those who are affected by organizational decisions are increasingly concerned about whether AI systems will treat them with care, respect, and empathy, appropriately capture the complexities and nuances of human behavior, and generate decision outcomes that are equitable and just – to name a few (cf. Binns et al., 2018; McGuire & De Cremer, 2022; Kellogg et al., 2020). In the present article, we broadly define AI as computational systems that act and decide in ways that *seem* intelligent and use the term "AI" as an umbrella term to capture both older classifications such as expert systems and linear models as well as more recent techniques such as machine learning and deep learning systems (Langer & Landers, 2021).

The predominant scholarly and industry approach to addressing these concerns about AI fairness has been a rational one. Rooted primarily in computer science, this approach adopts a technological perspective to define fairness in terms of probabilities and other mathematical terms (Corbett-Davies et al., 2017; Green & Chen, 2019; Verma & Rubin, 2018), thereby painting fairness as an objective process. However, humans perceive fairness *subjectively*, relying less on probabilities, and instead on intuition and affective information (De Cremer, 2007; Barsky & Kaplan, 2007). Fairness, in other words, is in the eye of the beholder.

Recent scholarship, especially in the field of Human-Computer Interaction (HCI), has been increasingly attentive to these perceptual aspects of AI fairness. A number of recent HCI studies have begun to examine how fairness is perceived by designers, end-users, decision-subjects, and various other stakeholders. Despite this growing interest, however, there remain a few critical problems in the current literature. First, although existing research has examined perceived fairness within specific topical domains such as learning analytics (cf. Hakami & Hernández Leo, 2020) and healthcare (cf. Rajkomar et al., 2018), as well as the implications of fairness perceptions for the technical design of AI systems (cf. Robert et al., 2020), what remains missing in the literature is a general integrative framework that analyzes fairness perceptions from the point of view of those who are affected by AI-augmented decisions.[1] Second, even though HCI scholars have long acknowledged the need to carefully consider the broader organizational context in which technologies are commonly developed and implemented (cf. Grudin, 1996; Kuutti & Bannon, 2014; Zhang et al., 2004), current research on perceived AI fairness has paid relatively little attention to such contextual factors (De Cremer, 2020).

As sites that provide the contextual background for most contemporary human-AI interactions, organizations play an

CONTACT Devesh Narayanan devesh.narayanan@nus.edu.sg Centre on AI Technology for Humankind, NUS Business School, National University of Singapore, 15 Kent Ridge Drive, 119245, Singapore

important role in shaping decision-subject perceptions of AI fairness. Moreover, the field of organizational justice has, over several decades, developed a rich body of conceptual frameworks and empirical findings that characterize the subjective fairness perceptions of those affected by organizational decisions (cf. Colquitt et al., 2013; Cropanzano et al., 2001; De Cremer & Tyler, 2005). As such, in our view, there is immense value in connecting this well-established organizational justice literature to the nascent yet burgeoning scholarship in AI fairness: to systematically categorize existing research on this topic, and in turn, to identify interesting opportunities for future research.[2]

To this end, in this paper, we conduct an integrative review[3] of existing empirical research on decision-subject perceptions of the fairness of AI-augmented decision-making processes. To organize and integrate this literature, we draw on a well-established framework in the organizational justice[4]literature – ie, Colquitt's (2001) dimensions of perceived fairness (procedural, distributive, interpersonal, and informational fairness). In addition, we also examine the impact of potential moderators for each dimension of perceived fairness and consider potential mechanisms that explain variations in fairness perceptions of AI-augmented decision-making. Finally, we conclude by discussing how fairness perceptions may change over time and, because research in this area is still in its early stages, we outline an agenda for potential future research in this domain. By synthesizing the existing literature on AI fairness perceptions into a validated framework of organizational justice and analyzing this phenomenon under varying conditions, this review advances our understanding of perceived AI fairness and offers new insights for both scholarly and industry discussions around this critical topic.

2. Artificial intelligence in organizations and perceptions of fairness

The significant uptrend in AI adoption in organizations has naturally given rise to concerns about its fairness (Kellogg et al., 2020). This is reflected in governance frameworks around the world that have positioned fairness as a core component of responsible, human-centered AI (cf. Butcher & Beridze, 2019; Jobin et al., 2019). Organizations have also gone to great lengths to commit to and promote policies on the fair development and use of AI (Bird et al., 2020; Walker & Croak, 2021). However, despite a recent explosion of such frameworks and policies, decision-subjects remain concerned about the fairness of AI deployments by organizations.

Within organizations, these concerns have been aimed at the use of AI systems to evaluate employee performance, hire new employees, promote existing employees, manage employee tasks, and allocate resources which frequently elicit negative fairness perceptions among employees (Acikgoz et al., 2020; Langer and Landers, 2021). For instance, employees have expressed concerns that AI systems may take too much decision-power (De Cremer & McGuire, 2022), as well as concerns about the inability of AI systems to appreciate the uniqueness of each employee (Longoni, Bonezzi, & Morewedge, 2019), to provide feedback on how decisions are made (Lee, 2018), and to provide sufficient transparency on its inner workings (Shin & Park, 2019). Moreover, as organizations increasingly use AI-augmented decision-making processes to deliver key goods and services, the fairness perceptions of external stakeholders (eg, current and potential customers) are also crucial to consider. For example, users of search engines have expressed concerns about biased and discriminatory search results (cf. Goldman, 2005; Noble, 2013); users of chatbots and other such AI-based service tools are concerned about the lack of interpersonal respect and care (cf. Barnett et al., 2021); and users of social media platforms are concerned about the algorithmic amplification of certain voices and the silencing of others (cf. Karizat et al., 2021; Swart, 2021), to name but a few.

Given such complexities, the currently predominant rationalistic approach to AI fairness – focused on documenting prescriptive rules and technical adjustments for designing AI systems that yield "objectively" fairer outcomes – seems incomplete. Rather, we also need a more descriptive perspective on AI fairness: grounded in firm evidence about how people *actually perceive* the fairness of using AI systems in organizational decision-making processes. Such an approach would therefore foreground *subjective* fairness perceptions, rather than focusing solely on objective metrics. Evidence for the success of such approaches – whereby a descriptive and subjective approach leads us to important conclusions that would have otherwise been inaccessible if only a prescriptive and rationalist approach were to be adopted – is clear in the fields of behavioral economics, behavioral finance, and behavioral operations, among others (Barberis & Thaler, 2003; Camerer, 1999; De Cremer et al., 2013; Gino & Pisano, 2008). For this reason, the present paper follows this second (subjective) path and reviews the existing empirical literature by making use of an established and well-researched theoretical framework on the distinctive dimensions of perceived fairness. This way, our review can bring to the fore important insights that can be used by organizations to better understand how to use AI in ways that are *perceived* as fair by decision-subjects.

2.1. Dimensions of perceived fairness

Organizational scholars have long paid careful attention to decision-subject (or, more narrowly, employee) perceptions of fairness (Greenberg, 1990, Colquitt et al., 2023). These fairness perceptions were deemed important to study because, when employees believe they are unfairly treated, this often results in a range of retaliatory or counterproductive behavior – such as stealing from the organization, being hostile to colleagues, and even resigning from jobs (Daileyl & Kirk, 1992; Dietz et al., 2003; Greenberg, 1993). As such, scholars quickly sought to develop a theoretical framework that comprehensively breaks down the different types of fairness concerns that employees have regarding decisions made in organizations (Colquitt, 2001). Examining the issue of fairness perceptions of AI according to Colquitt's (2001) four dimensions of organizational justice is particularly apt because AI presents unique challenges and opportunities that relate to each dimension: procedural, distributive, interpersonal, and informational. By considering the definition of each dimension and how it relates to the unique characteristics of AI, the natural fit of integrating AI fairness research according to these dimensions becomes clear.

Organizational justice scholars initially examined the issue of *distributive fairness*, ie, the fairness of *decision outcomes* received (Adams, 1965; Deutsch, 1975; Leventhal, 1976). Decision outcomes that are consistent with implicit or explicit norms for allocating resources are generally perceived as high on distributive fairness (Colquitt, 2001). Such allocation norms tend to fall into two broad categories: equity and equality. Norms of equity are concerned with allocations being proportionate to the effort, a person's contributions, and a sense of deservingness (Deutsch, 1975). Norms of equality, by contrast, simply suggest that resources should be allocated equally among all groups of people (Cook & Hegtvedt, 1983). Although both norms have been researched extensively in the organizational justice literature, today, scholars usually appeal to norms of equity when discussing distributive fairness. In the case of AI making allocation decisions, employees may worry about AI systems violating norms of equity since they are incapable of truly appraising an employee's deservingness, merit, and contribution (cf. Binns et al., 2018). On the other hand, the consistency with which AI applies decision rules, in combination with the possibility of making these rules transparent, could potentially bolster perceptions of distributive fairness.

Subsequent work gradually started to recognize that when allocating outcomes, fairness perceptions were also influenced by receivers' attention to how those outcomes were allocated, which moved the research focus to the fairness of the decision-making process, which is referred to as *procedural justice* (Leventhal, 1976; Thibaut & Walker, 1978;). Decision procedures are considered to be fair when they are free from bias, correctible, accurate, and provide people with the opportunity to influence decision outcomes, such as through voicing concerns (cf. Lind & Tyler, 1988). The use of AI in decision-making, for instance, may raise concerns over whether people will be able to influence decision outcomes, depending on the extent to which decision-making is fully automated (Vimalkumar et al., 2021). In addition, the opaque nature of more complex AI models – such as those based on deep learning techniques – may also call into question the correctability of such decision processes (Castelvecchi, 2016). Both examples illustrate how AI poses unique challenges when procedural fairness perceptions are concerned.

With the move in focus to also paying attention to process fairness in the decision-making stage, it became clear that attention to decision-making processes was also motivated by people's concerns about the interpersonal treatment they receive when decisions are made (Bies & Moag, 1986). *Interpersonal justice* is the degree to which people are treated with dignity and respect (Tyler & Bies, 2015), and importantly builds upon procedural justice as it includes whether sensitivity and due care are present in how decision processes are carried out and how decision outcomes are communicated (Greenberg, 1990). Whether AI can serve as a comparable substitute to human decision-makers in facilitating interpersonal justice is of major interest in the domain of AI fairness perceptions (Formosa et al., 2022; Latonero, 2018), as the capacity to demonstrate sensitivity and treat others with respect and dignity is frequently cited as a uniquely human trait (De Cremer et al., 2022).

Informational justice is determined by the quality with which explanations are provided regarding how decisions are made, and why certain outcomes were reached (Greenberg & Cropanzano, 1993). Explanations are considered more just when they are timely, reasonable, and specific (Shapiro et al., 1994). Much emphasis has been placed on the need for explainability in AI (Gunning et al., 2019) so that decision-subjects can understand how these decisions are made and ultimately determine whether they are just (Barredo Arrieta et al., 2020). While AI is positioned to give explanations instantaneously, it is unclear whether such explanations are regarded as appropriate or specific (Adadi & Berrada, 2018), and hence informationally fair.

These different categories of fairness perceptions were combined into Colquitt's (2001) framework, which thus offers clear and well-defined categories that humans use when evaluating the fairness of a decision-maker. In a similar vein, we, therefore, argue that AI as a decision-maker will also be subjected to these distinctive ways in which users perceive the fairness of how AI makes decisions and the outcomes these processes reveal. Moreover, from a conceptual point of view, organizing existing empirical research on AI-based decision-making along the four fairness dimensions also allows us to draw upon decades of prior research on organizational justice to explicate the strengths, weaknesses, and opportunities for future research on perceived AI fairness. For these reasons, we consider Colquitt's (2001) framework as especially well-suited to serve as a theoretical anchor for our review.[5]

3. Literature review methodology

Our review corpus comprises academic publications – ie, articles in peer-reviewed journals and conference proceedings – that explicitly focus on perceptual dimensions of AI fairness, published between 2001 and 2022. Such papers can be found in the literatures of computer science, human-computer interaction, management, psychology, and various other social science fields – and as such, our review was deliberately field-agnostic. To ensure that we could capture this breadth of scholarship, we consulted three multidisciplinary search databases (Web of Science, Scopus, and Google Scholar) as well as several relevant discipline-specific databases (including ArXiv, PsyrXiv, Business Source Premier, and SSRN). We queried each of these databases with a three-part search string, comprising: (a) keywords related to AI systems (ie, "algorithm*", "machine learning", "artificial intelligence", "robot*"), (b) keywords related to fairness (ie, "justice", "fair*", "unfair*")), and (c) keywords related to perceptions (ie, "judg(e)ment", "perceive*", "perception"). After deleting duplicates, we manually screened the titles, abstracts, and introductions of the papers in our search results, to exclude papers that focused on (a) technical/mathematical notions of AI fairness without reference to human perceptions, and (b) justice/fairness perceptions in non-AI contexts. At the end of this screening process, our final review corpus comprised 144 unique papers. Based on this approach, we were already able to make a first interesting observation, which is that a majority of these papers (approximately 70%) were published since 2019, evidencing the timeliness and rapidly growing importance of this research area.

The first four authors then proceeded to analyze each paper in our corpus, proceeding in two stages. In the first stage, each author was randomly assigned an average of 35 papers to categorize according to the four dimensions of perceived fairness (distributive, procedural, informational, and interpersonal). We also summarized the main findings of each paper, the social/task context (eg, recruitment in organizations, work scheduling, content moderation, etc.), and the type of AI involvement in decision-making (eg, human vs. AI decision-maker; AI + human oversight; AI decision-maker only). These categorizations were explicit in some papers, but in the remaining cases, a subjective evaluation for categorizing was required. As such, to ensure that our categorizations were consistent, the authors met to discuss and achieve consensus about papers that were difficult to categorize. At the end of this process, approximately half of these papers ($n = 71$) were deemed to contain empirical findings that were directly connected to one or multiple fairness dimensions.

In the second stage, each author was assigned to one of the four fairness dimensions and reanalyzed each paper categorized under their respective dimension. In this stage, we focused on (a) synthesizing findings across each of the four dimensions of perceived fairness, (b) identifying common themes and proposed mechanisms for each dimension, and (c) identifying gaps for future research. We engaged in several rounds of critical and analytical discussion to synthesize overall findings, explain our reasoning to each other, and address any remaining ambiguities and discrepancies. In this way, our two-stage qualitative review process, interspersed with frequent meetings and discussions, ensured that each of the 71 categorized papers were close-read independently by at least three authors and that all four authors were closely involved in deriving, discussing, and evaluating general insights related to our review topic.

After the first stage, however, 73 papers could not be included for our integrative review as they were conceptual papers, commentaries, and/or reviews that did not contain new empirical findings. Even so, these papers provided useful context and analysis to identify strengths and weaknesses in existing research on the perceptions of AI fairness, and therefore, where to locate the best opportunities for future research. As such, when writing up this review – especially in the *General Discussion* section – all five authors reviewed and discussed our notes, summaries, and reflections pertaining to these 73 papers, to ensure that, wherever appropriate, our review also included relevant perspectives from the literature that fell outside the specific focus of our analysis. A complete list of the articles included in our review corpus may be found in Table 1.

4. Theoretical review and analysis

As such, by qualitatively analyzing and coding the papers in our corpus, we were able to map the existing literature to the dimensions of the organizational justice framework (ie, procedural, interpersonal, informational, and distributive). In so doing, we found that a range of factors moderate and mediate the effects of AI decision-making along the different dimensions of fairness perceptions. These findings are presented below.

4.1. Procedural fairness

A considerable portion of research on perceptions of fairness as it relates to AI has been in the context of procedural fairness. However, this research does not point to a clear relationship between AI decision-making processes and procedural fairness; indeed, as we discuss below, some research finds that AI processes are considered less fair than comparable human processes, and other research finds that AI processes are perceived as more fair than comparable human processes. For example, people believed that moderation of online political content was less fair when done via AI than when done via a person (Wojcieszak et al., 2021); however, warehouse workers perceived the allocation of task assignments to be more procedurally fair via AI than via a human (Bai et al., 2021).

One explanation for these discrepancies is that it depends on the task at hand. For example, there may be a reason to believe that AI procedures are perceived as less fair than humans when the task is expected to require uniquely human skills (eg, requiring subjective judgment and/or emotional capabilities) rather than mechanical skills (eg, processing quantitative data for objective assessments; Lee, 2018). For example, numerous research studies showing that people perceive AI to be less fair than humans can be found in the context of job candidate selection (Acikgoz et al., 2020; Dineen et al., 2004) and recruitment interviews (Folger et al., 2022; Langer et al., 2019; Nørskov et al., 2020). These contexts often require making subjective assessments of candidates' prior experiences, capabilities, as well as their interpersonal qualities via an interview. The human skills necessitated by these tasks thus strongly influence perceptions of AI fairness. Additionally, "human" tasks like candidate selection and interviewing may also be prone to discrimination, and algorithmic unfairness activates concerns about racial injustice and economic inequality (Woodruff et al., 2018) that cannot be easily controlled for (see eg, De Cremer & De Schutter, 2021). This explanation can help make sense of the two seemingly contradictory findings cited at the beginning of this section. Moderation of political content online, the topic of Wojcieszak et al. (2021)

Table 1. Summary of articles included in our review corpus.

	Procedural	Distributive	Informational	Interpersonal
Acikgoz et al. (2020)	x			x
Ahnert et al. (2021)	x			x
Albach & Wright (2021)	x	x		
Araujo et al. (2020)		x		
Bai et al. (2021)	x	x		
Bankins et al. (2022)	x			x
Barlas et al. (2019)		x		
Binns et al. (2018)	x	x	x	
Brown et al. (2019)	x			x
Chang et al. (2021)		x		
Chang et al. (2021)	x			
Chen et al. (2021)	x			
Cheng et al. (2021)		x		
Choi et al. (2021)			x	x
De Cremer and Chun (2021)	x			x
Dineen et al. (2004)	x			
Dodge et al. (2019)			x	
Eslami et al. (2019)			x	
Ferraro et al. (2021)	x		x	
Fleiß et al. (2020)			x	
Folger et al. (2022)	x			
Gamez et al. (2020)				x
Gonçalves et al. (2021)		x		
Grgić-Hlača et al. (2018a)	x			x
Grgić-Hlača et al. (2018b)	x	x		
Gupta et al. (2021)	x	x		
Harrison et al. (2020)		x		
Helberger et al. (2020)	x			
Hobson et al. (2021)	x			x
Htun et al. (2021)		x		
Kaibel et al. (2019)	x			x
Kasinidou et al. (2021)		x		
Kieslich et al. (2022)	x			
Lai et al. (2020)		x		
Langer et al. (2019)	x			
Langer et al. (2021)	x		x	
Lee (2018)	x			x
Lee and Baykal (2017)	x	x		
Lee and Rich (2021)	x			x
Lee et al. (2019a)	x			
Lee et al. (2019b)	x	x		
Marcinkowski et al. (2020)	x	x		
Miller and Keiser (2021)	x			
Mirowska and Mesnet (2022)	x	x	x	x
Nagtegaal (2021)	x			
Newman et al. (2020)	x		x	
Noble et al. (2021)	x			x
Nørskov et al. (2020)	x			x
Ogunniye et al. (2021)			x	
Ötting and Maier (2018)	x			
Pierson (2017)		x		
Rader et al. (2018)			x	
Saha et al. (2020)		x		
Saxena et al. (2020)		x		
Schoeffer et al. (2021)			x	
Schlicker et al. (2021)		x	x	x
Shulner-Tal et al. (2022)			x	
Skinner et al. (2020)			x	
Smith et al. (2020)		x	x	
Sonboli et al. (2021)			x	
Srivastava et al. (2019)		x		
Suen et al. (2019)	x			
Tomaino et al. (2020)			x	
Vaccaro et al. (2020)	x			
van Berkel et al. (2019)	x			
van Berkel et al. (2021)			x	
Wang (2018)	x		x	
Wang et al. (2020)		x		
Wangmo et al. (2019)	x	x		x
Wonseok et al. (2021)		x		

(*continued*)

Table 1. Continued.

	Procedural	Distributive	Informational	Interpersonal
Woodruff et al. (2018)	x			
Zhang and Yencha (2022)		x		

Articles in our corpus that could not be categorized according to the four fairness dimensions: Adadi and Berrada (2018); Alarie et al. (2018); Antifakos et al. (2005); Barredo Arrieta et al. (2020); Banks (2021); Bansal et al. (2021); Binns (2022a); Binns (2022b); Chouldechova et al. (2018); Cowgill (2018); de Fine Licht and de Fine Licht (2020); DeVito et al. (2017); Diakopoulos and Koliska (2017); Diakopoulos (2015); Dietvorst et al. (2018); Edwards and Veale (2017); Firestone (2020); Goldfarb and Lindsay (2020); Goodman and Flaxman (2017); Green (2020); Grgić-Hlača et al. (2020); Hannan et al. (2021); Höddinghaus et al. (2021); Holstein et al. (2019); Langer et al. (2022); Lepri et al. (2018); Li et al. (2021); Litoiu et al. (2015); Logg et al. (2019); Lyons et al. (2021); Martin (2019); Masrour et al. (2020); Mehrabi et al. (2022); Miller (2019); Mittelstadt et al. (2016); Rader and Gray (2015); Schoeffer et al. (2021); Selbst et al. (2019); Shandilya et al. (2021); Shin (2020); Shin and Park (2019); Shin et al. (2020); Short et al. (2010); Simshaw (2018); Skewes et al. (2019); Stai et al. (2020); Starke et al. (2021); Stellmach and Lindner (2019); Tulk and Wiese (2018); Vaccaro and Waldo (2019); van Berkel et al. (2022); Wang and Yin (2021); Werth (2019); Zahedi et al. (2020); Elahi et al. (2021); Shin (2022); Park et al. (2021); Schadenberg et al. (2021); Telkamp and Anderson (2022); Mitchell et al. (2021); Hunkenschroer and Luetge (2022); Kushwaha et al. (2021); Köbis and Mossink (2021); Lima et al. (2021); Schick and Fischer (2021); Dietvorst and Bartels (2022); Wiener et al. (2021); Wojcieszak et al. (2021); Charisi et al. (2021); Kleanthous et al. (2022); Morse et al. (2022); Schoeffer and Kuehl (2021); Zhou et al. (2021).

research, is likely viewed as fundamentally human as politics are notoriously subjective (Van Bavel & Pereira, 2018) and as such may be perceived as requiring a human touch. By contrast, task assignments for warehouse workers, as studied by Bai et al. (2021), may be viewed as more rational and calculative, and therefore, more acceptable for a machine to do.

The relationship between the enactment of decision-making procedures by AI versus humans and the perceived fairness of this enactment may also be unclear because it depends on task *complexity*. Multiple studies have shown that for tasks that are high on complexity—for example, ones that involve multiple intertwined factors or multiple phases—AI enacting procedures are perceived as less fair than humans' enactment (Gupta et al., 2021), whereas for simpler tasks, AI enacting decision-making procedures are perceived as more fair (Nagtegaal, 2021). This may be because for complex tasks, people believe that algorithms are reductionistic, removing important context in an effort to quantify the decision to arrive at an objective answer (Newman et al., 2020).

Moreover, fairness perceptions of AI-based decision-making may also depend on how decision-subjects view the (un)fairness of existing decision-making processes prior to the adoption of AI. As Stapleton et al. (2022)[6] find in their study of prediction systems for child welfare support: in contexts where current (human-driven) decision-making processes are known to be fraught with biases and discrimination, people worry that the use of AI might further entrench and exacerbate the unfair treatment of minoritized groups.

Still, other research suggests that the role of AI in procedural fairness remains unclear. For example, Suen et al. (2019) found that the presence of AI did not influence perceptions of fairness, specifically in an interview setting. Other research found no effect of AI on procedural fairness in the context of task allocation and opportunities for work training (Ötting & Maier, 2018). These findings signal opportunities for future research to disentangle the factors by which AI may or may not influence procedural fairness perceptions of the decision-subject.

4.2. Interpersonal fairness

Of the four dimensions of fairness, interpersonal fairness was the one dimension where there was a broad consensus on the findings reported: ie, that AI-driven decision-making processes were generally perceived to be less interpersonally fair. However, it is important to note that a majority of the studies in our corpus focused on interpersonal fairness from the perspective of human resource management (HRM)-related decisions or the various stages of employee recruitment, such as the initial screening stage or the subsequent interview stage (eg, Acikgoz et al., 2020; Bankins et al., 2022; Kaibel et al., 2019; Noble et al., 2021; Nørskov et al., 2020). It is therefore possible that our general finding only applies to this specific HRM context, and as such, more research is needed to carefully examine how perceptions of interpersonal fairness vary across contexts.

Further investigations into *why* algorithmic decision processes are generally perceived to be less interpersonally fair indicate that, when it comes to a job candidate's intentions to pursue a role based on how the company conducts its interviews, a lack of two-way channels of communication, as is the case with AI-based interviews, may send a negative signal to the applicant that the company does not care about its employees (Acikgoz et al., 2020). In other words, an interview conducted by AI may make job candidates feel as though they are just a number and not a unique individual, as the company did not even take the time to conduct interviews in person. This preference for human interaction is echoed by other research papers that examine interpersonal fairness in contexts other than HRM as well. For instance, when it comes to healthcare scheduling, perceptions of interpersonal fairness were stronger when a human made the scheduling decision, compared to when an automated agent made the same decision (Schlicker et al., 2021).

Other studies that consider perceptions of interpersonal fairness in determining algorithmic fairness indicate that there are certain contexts where the "human touch" is critical, and an algorithm alone cannot accomplish what a human can. For instance, when it comes to the child welfare system, decision-subjects perceive algorithmically driven decision-making processes to be less interpersonally fair compared to human-driven (Brown et al., 2019). Similarly, when it comes to the use of intelligent technologies to care for the elderly and disabled, these technologies are perceived to be less interpersonally fair because decision-subjects care about human contact and empathy and deem these to be essential for care that is effective and morally acceptable (Wangmo et al., 2019).

In considering additional reasons for why algorithmic decision processes may generally be perceived to be less interpersonally fair compared to human ones, it makes sense to consider what makes a human, a human. One specific human characteristic is our ability to apologize for a misdemeanor. Choi et al. (2021) tested apologies as a recovery tactic to increase interpersonal fairness. They found that following a process service failure, an apology allows the wronged to feel warmth and in turn, satisfaction, toward the robot perpetrator. They note that this is the case especially when the apology is delivered by a humanoid, rather than a nonhumanoid service robot.

While all of this explains how and why algorithmic decision-making is generally perceived to be less interpersonally fair compared to human decision-making, it is also important to consider potential moderators to this relationship. One of these moderators is the outcome of the decision. Bankins et al. (2022) found that algorithmically driven decision-making processes are generally perceived to be less interpersonally fair compared to a human decision-maker, even when the AI decision-maker metes out a positive decision. However, when the decision outcome is unfavorable, no differences were detected across the AI and human conditions. Other research reports similar findings although they do not examine fairness perceptions but rather "favorable reactions" (Yalcin et al., 2022).

Another moderator to this relationship between algorithmic and human-driven decision-making processes and interpersonal fairness is *trust* perceptions. Within the context of medical AI, although one might expect patients with higher levels of medical mistrust to be more accepting of algorithmic technologies, findings indicate that patients with high medical mistrust perceive the algorithmic decision to be just as interpersonally fair, or rather unfair, as a human decision (Lee & Rich, 2021). In contrast, those with low levels of mistrust in the medical system perceive the algorithmic decision as less interpersonally fair than human decisions (Lee & Rich, 2021). Future research should examine how perceived trust – in relation to both AI systems themselves as well as the institutions that deploy these systems – might influence perceptions of interpersonal fairness. Previous research in the management literature has found that trust perceptions can sometimes act as a substitute for fairness perceptions, especially in cases where fairness information is not readily available (cf. van den Bos et al., 1998). It would be interesting to see if similar connections between trust and fairness can also be found in the context of AI-based decision-making.

4.3. Informational fairness

In our review of the literature, we found that relatively little research attention has been paid to examining decision-subject perceptions of informational fairness in relation to AI-driven decision-making. This is somewhat surprising since concerns about the "explainability" and "transparency" of AI systems have long dominated research and policy-making conversations around what makes for an ethical AI (cf. Fjeld et al., 2020; Jobin et al., 2019). However much of this extant research on explainable and transparent AI has a largely technical focus on which types of technology features might make black-boxed AI systems explainable and transparent. Moreover, empirical research on how decision-subjects interact with explainable AI systems tends to focus on (a) how decision-makers can use the explanations provided to improve their own decision-making practices, and (b) how decision-subjects can use explanations to understand whether the decisions that affect them are distributionally or procedurally fair (cf. Green & Chen, 2019; Hase & Bansal, 2020; Kaur et al., 2020; Yang et al., 2021). As such, little research so far has specifically examined how the provision of information about AI-based decision-making processes affects decision-subject perceptions of informational fairness.

A small number of papers in our corpus ($n = 4$) focused directly on comparing perceived informational fairness in human- versus AI-driven decision-making contexts, and a few additional papers ($n = 16$) more generally examined how the provision of different kinds of information by AI and human decision-makers affect perceptions of informational fairness. Although it is difficult to make conclusive determinations from this relatively small sample, we find that, in general, AI involvement in decision-making processes tends to adversely impact perceptions of informational fairness. Several articles noticed an interesting asymmetry: increasing the transparency of decision-making processes tends to lead to greater perceived informational fairness when humans were the decision-makers, but not when AI was the decision-maker (Newman et al., 2020; Rader et al., 2018; Schlicker et al., 2021). As such, even in cases where people might perceive AI decision-makers as more "transparent" than humans (for instance, when technical "explainability" features are used to demonstrate the inner logic of an AI system, cf. Gonçalves et al., 2021), this does not necessarily mean that they, therefore, perceive AI-driven decision-making processes as informationally fairer.

Prior research on informational fairness in the organizational justice literature has suggested that for explanations to be perceived as fair, they must also be recognized as reasonable, understandable, and responsive to the decision-subject's needs and concerns (Greenberg & Cropanzano, 1993). Our reviewed papers largely align with this principle. Explanations from AI decision-makers were found to improve perceptions of informational fairness when they articulated the underlying moral purpose or values behind the decision rather than simply describing the inner logic and procedures followed by the AI decision-maker (Tomaino et al., 2020). In turn, Dodge (2019) and Shulner-Tal et al. (2022) find that explanations that provide information about the AI-system decision-making logic are perceived as more informationally fair than those that provide comparative information about disparities in outcomes between different decision-subjects (eg, how an affected decision-subject fared in comparison with different demographic groups). In other words, moral- or value-based explanations seem to be the most related to positive perceptions of informational fairness, followed by procedural explanations, and finally by outcome-disparity explanations. Finally, the manner in which complex information is presented to the decision-subject also makes a difference to informational fairness perceptions. Specifically, complex visual explanations in the forms of scatterplots and graphs lead to diminished fairness perceptions as compared to simple text-based visualizations (Schoeffer et al., 2021; Van Berkel et al., 2021).

Several studies also reported that perceptions of informational fairness were not only affected by the content and structure of the provided information itself but also by broader contextual factors. When AI-driven decisions were favorable (versus unfavorable) to the decision-subject, AI explanations had a positive effect on perceived informational fairness (Bankins et al., 2022; Shulner-Tal et al., 2022; Yalcin et al., 2022). Relatedly, Eslami (2019) find that the level of engagement with, and personal gain from, AI decision-makers are an important moderating factor: ie, decision-subjects would be more likely to perceive explanations from AI decision-makers that they engage heavily with, and/or gain a lot from, as informationally fair. Finally, Schoeffer et al. (2021) find that the "actionability" of recommendations is another key factor: decision-subjects are more likely to perceive explanations as informationally fair if they can see an actionable path to using these explanations to challenge unfair outcomes and processes.

4.4. Distributive fairness

Our literature review revealed that there is much work to be done in understanding how decision-subjects perceive the distributive fairness of AI-augmented decision-making. Despite the fact that a sizeable number of papers in our corpus examined perceptions of distributive fairness in some sense, the number of articles that *directly* compare AI vs. human decision-making along this dimension is modest ($n = 7$) and as such, comparative trends and patterns for this fairness dimension are difficult to discern. Evidence can be found on both sides as to whether perceptions of distributive fairness are greater or lower for decisions made by AI (vs. humans). For instance, in the context of university admission decisions, the outcomes of these decisions were considered fairer when AI made those decisions, as opposed to when a human committee did (Marcinkowski, Kieslich, et al., 2020). In a similar vein, people were found to exhibit more tolerance towards an AI system that made decisions that resulted in unequal outcomes, relative to when a trained human expert (a psychologist) made the same decisions (Wang, 2018). The allocation of tasks to warehouse workers was also found to be fairer when administered by an algorithm as opposed to a human (Bai et al., 2021). On the other hand, distributive fairness perceptions were greater when humans (vs. AI) were responsible for tagging/labeling images for dating profiles (Barlas et al., 2019), and for making decisions as an umpire in baseball (Wonseok et al., 2021). Interestingly, this latter effect found that when the AI umpire was anthropomorphized, participants considered it to be just as fair as human umpires. However, some studies also found there to be no differences in perceptions of distributive fairness between human and AI decision agents across legal, media, and healthcare contexts (Araujo et al., 2020; Schlicker et al., 2021). This suggests that there may be a number of boundary conditions – such as task objectivity, and anthropomorphizing – which attenuate negative perceptions towards AI decision agents that warrant further investigation, given the nascent and preliminary nature of the investigations thus far.

There are, however, a few articles that potentially elucidate a number of moderating factors that attenuate or exacerbate perceptions of distributive fairness of AI. Firstly, Gupta et al. (2021) found that perceptions of distributive fairness of AI-based decisions were lower when the decision task was relatively more complex. In addition, this negative effect of task complexity was exacerbated when transparency in how that decision was made is also low. Additionally, in the domain of AI-generated music recommendations, people who reported high levels of openness to experience were relatively less likely to express concerns about the fairness of the AI recommendation system (Htun et al., 2021). Alternatively, people who reported high levels of conscientiousness were more likely to express fairness concerns. A further factor that alters perceptions of distributive fairness is the degree to which decision-subjects can participate in determining the factors that determine how AI decisions are made, whereby increased opportunities for participation promoted the perceived fairness of decision outcomes (Lee et al., 2019b). Collectively, these findings highlight that perceived effort, task complexity, decision transparency, individual differences in personality, and degree of human involvement may all alter distributive fairness perceptions of AI.

Consistent with theories of distributive fairness (Cook & Hegtvedt, 1983; Leventhal, 1976), support was also found in the literature review for the existence of both equity (Kasinidou et al., 2021; Lee & Baykal, 2017; Mirowska & Mesnet, 2022; Saxena et al., 2020) and equality (Cheng et al., 2021; Harrison et al., 2020; Srivastava et al., 2019; Wangmo et al., 2019) as central norms when determining the distributive fairness of decisions made by AI. In support of equity as a fair distribution norm, "proportional equality" was the preferred AI allocation strategy for the division of rent, house, credit, etc., (Lee & Baykal, 2017). Saxena et al. (2020) compared three AI allocation strategies - (1) treat similar individuals similarly, (2) never favor a worse individual over a better one, and (3) calibrated fairness – and found that calibrated fairness was the most preferred strategy. The preference for equity as a guiding distribution norm in decisions made by AI was also found in the context of job interviews (Mirowska & Mesnet, 2022; Saha et al., 2020). Calibrated fairness was defined as a strategy that

Table 2. Key trends and overall findings from our review.

Procedural fairness	**No. of Papers**: 41 **Key Trends:** ● Decision-making tasks that require uniquely human skills (eg, subjective judgment, emotional capabilities, relational care, etc.) tend to be perceived as less procedurally fair when AI systems are involved. ● Complex decision-making tasks (eg, with multiple intertwined factors or multiple phases) are perceived as less procedurally fair when AI systems are involved.
Interpersonal fairness	**No. of Papers**: 17 **Key Trends:** ● In general, AI-augmented decision making is perceived to be less interpersonally fair. ● Decision-making contexts where two-way communication channels between the decision-maker and decision-subject are desirable (eg, job interviews) are perceived as less interpersonally fair when AI systems are involved. ● In contexts where users are generally mistrustful of (human-driven) decision-making, AI involvement does not seem to affect perceptions of interpersonal fairness.
Informational fairness	**No. of Papers**: 20 **Key Trends:** ● Increasing the transparency of decision-making processes tends to lead to greater perceived informational fairness when humans are the decision-makers, but not when AI systems are involved. ● When explanations in AI-driven decision-making are recognized as reasonable, understandable, and responsive to the users' needs and concerns, they are more likely to be perceived as informationally fair.
Distributional fairness	**No. of Papers**: 27 **Key Trends:** ● General trends in the perceived distributive fairness of AI-augmented decisions are hard to discern, and more careful, systematic research on this topic is needed. ● Greater task complexity, lower decision transparency, individual differences in personality (ie, lower openness, higher conscientiousness), and fewer opportunities to participate in shaping decision-outcomes were all found to adversely affect perceptions of distributive fairness in AI-augmented decision-making.

"selects individuals in proportion to their merit" (Saxena et al., 2020, p. 3). On the other hand, a number of papers also highlighted equality as an important factor. Across AI predictions of criminal risk of reoffending, recommendation systems, skin cancer risk, flu risk, and risk of child maltreatment, people found the predictions to be most fair when they equalized the probability of false positives and true positives across all demographic groups (Cheng et al., 2021; Harrison et al., 2020; Pierson, 2017; Smith et al., 2020; Srivastava et al., 2019). Therefore, both notions of equity and equality emerge as critically relevant when perceptions of distributive fairness are formed in AI decision-making.

4.5. Spillover effects

One interesting point of investigation for this review is whether there is any evidence for perceptions of fairness dimensions positively influencing one another. Although there were few examples of this in the literature, the findings are still somewhat encouraging. In both the work of Grgić-Hlača et al. (2018b) and Binns et al. (2018), it was found that when fairness perceptions of the decision-making process are high (procedural fairness), this in turn also generally promoted fairness perceptions of decision outcome (distributive fairness). These findings indicate that decision processes are an important area for organizations and developers to initially focus on as they produce positive spillover effects. Although evidence for such spillover effects is sparse, we believe that there might be interesting opportunities for future researchers to examine when and how fairness perceptions across the four different dimensions can positively build upon each other, in the context of AI-based decision-making.

5. General discussion

This review serves to collate and integrate the growing amount of research on AI fairness as perceived by decision subjects and connect this research to a mature body of scholarship on organizational justice. To do so, we categorized our papers according to Colquitt's (2001) four dimensions of fairness and examined how fairness perceptions of AI-augmented decision-making vary within and across these four dimensions. The key trends and findings of our review are summarized in Table 2.

Keeping these broad findings in mind, we can begin to outline an agenda for future research.

5.1. An agenda for future research

Although our review indicates that a fair amount of work has been done on the perceived fairness of AI decision-making, it also makes clear that much work still remains to be done. To better understand what future research on this topic might look like, it is instructive to turn once again to the organizational justice literature – which has, over several decades, carefully examined various factors that shape perceptions of fairness across the four dimensions, and how these fairness dimensions interact with one another.

Perhaps the most notable interaction effect observed in the organizational justice literature is between distributive and procedural fairness: ie, when procedural fairness is perceived to be low, people are more sensitive and react more violently to unfavourable outcomes, but not so when procedural fairness is perceived to be high (cf. Brockner et al., 1994), and this effect has been shown to hold across a variety of situations (cf. De Cremer et al., 2010). Relatedly, organizational justice research indicates that whether the outcome is positive or negative plays an important role, over and above perceptions of procedural fairness, in determining overall fairness perceptions (Daly & Tripp, 1996; Törnblom & Vermunt, 1999). Will such interactions between procedural and distributive fairness perceptions also be observed when the decision-maker is an algorithm? For now, there is evidence that suggests that decision-subjects are more likely to deem an algorithm's outcomes to be fair, if it rules in their favor, even if the algorithm is described as one that is procedurally biased against certain demographic groups (Wang et al., 2020). This would map onto the findings in the organizational justice literature. On the flip side, the interaction between procedural and distributive fairness may play out differently in the case of algorithms, as research findings indicate that people discount algorithms more so than humans when an error is made (Renier et al., 2021). This means that if a procedural error and/or an outcome error is made by an AI vs. a human, human managers may be able to get away with the error at rates greater than the AI, even if the AI performs at a level better than the human. More careful research is needed to carefully explore these interaction effects in the context of AI-augmented decision-making.

Organizational justice research also notes that interpersonal fairness often interacts with informational fairness, especially in environments that are deemed uncertain to the decision-subject (Brockner et al., 2021). In these scenarios, the provision of explanations is key to improving both informational and interpersonal fairness perceptions (Rolland & Steiner, 2007). Since interactions with AI systems are prone to be perceived as uncertain and ambiguous by humans (Dietvorst & Bharti, 2020), these findings should be especially relevant. However, as we noted above, our review found an interesting asymmetry between the provision of explanations and perceived informational fairness: ie, that explanations enhanced perceived informational fairness for human decision-makers, but not in AI-augmented decision-making contexts (cf. Newman et al., 2020; Schlicker et al., 2021). It would be interesting for future researchers to examine whether similar asymmetries also emerge in perceptions of interpersonal fairness in AI-augmented decision-making contexts, and in turn, carefully unpack the interactions between these two fairness dimensions.

With respect to distributive fairness, future research should address whether there are differences in perceptions of distributive fairness when the party affected is an individual versus a group, or even society as a whole. As AI increasingly takes on more managerial decision-making responsibilities, and as "algorithmic management" increasingly becomes a reality for many organizations (c.f. De Cremer, 2020; Jarrahi et al., 2021; Lee 2018), algorithmic "managers" will also make sensitive and morally significant allocation decisions for the groups of employees that they manage. Moreover, as computer scientists have argued, training machine learning algorithms to produce fair outcomes for individuals can sometimes come into conflict with producing fair outcomes for groups (Binns, 2020; Fleisher, 2021). Broadly speaking, this occurs because definitions of individual fairness may at times be mathematically incompatible with definitions of group fairness. Even though such technical considerations are beyond the scope of this review, the point to note here is that emphasizing individual fairness may diminish group fairness, and vice versa. In turn, this is bound to affect perceptions of the algorithm's distributive and overall fairness. If this is the case, can explanations, an element of informational fairness, be utilized to improve perceptions of overall fairness?

Another question that needs to be addressed is how perceptions of AI fairness differ cross-culturally. Organizational justice research finds that Americans and Japanese are more concerned with interpersonal fairness and less so with distributive fairness, as compared to Chinese and Koreans (Kim & Leung, 2007). Do these perceptions hold true even when the allocations are being made by a machine? Current research findings suggest that cultural variations indeed affect perceptions of AI fairness in a way that may be similar to perceptions of fairness in general. Gupta et al. (2021) find that when AI-based recommendations or outcomes are racially or gender biased, individuals from cultures that espouse collectivism, masculinity, and uncertainty avoidance, as is the case in China and Korea, are more likely to question the outcome bias. However, as van Berkel et al.'s (2023) point out in their comprehensive review, a large proportion of existing studies on AI fairness have thus far focused on North American samples – evidencing that much work remains to be done in understanding how AI fairness perceptions vary across cultures.

Taken together, by setting out the questions that we have identified here, we hope to provide input to developing a future research agenda that will enable the scholarly community to close the gap between research on fairness in the field of human-AI interaction and research on fairness in the field of organizational justice.

5.2. Fairness perceptions across time

As organizations increasingly develop and deploy AI systems for a variety of decision-making tasks, over time, decision-subjects will likely increase their exposure to AI-augmented decision-making. In turn, their attitudes and ideas about AI may also evolve over time. Indeed, although there is currently a sense of skepticism associated with allocation-based decisions made by AI, but given the fact that psychological commitments and perceptions change over time, it is possible that certain dimensions of AI fairness are particularly likely to be either well or ill-received over time - depending on the evolving needs and concerns of decision-subjects. As such, we propose that AI fairness may well be a function of an interaction effect between the opportunities that AI currently brings to the table and the ways human decision-subjects accommodate and collaborate over time with AI, consequently affecting their fairness perceptions (cf. De Cremer & Kasparov, 2021).

Prior organizational research on perceptions of fairness has suggested important changes in its trajectory over time (Jones & Skarlicki, 2013; Konradt et al., 2016; Streicher et al., 2012). Numerous factors, such as whether an entity's behavior violates expectations based on prior experience, can influence perceived fairness over repeated interactions (Jones & Skarlicki, 2013). From our overview of the state of the current literature, we could find no research directly testing how fairness perceptions develop and are shaped by repeated interactions between humans and AI. As such, this remains a crucial gap in our current understanding that warrants future research, especially since perceptions of technological agents are not exclusive to one-shot interactions and may indeed pervade organizations and other contexts of human life over time. To approach how fairness perceptions of AI might change in trajectory over time, we partially rely on a similar analysis conducted by Glikson and Woolley (2020), who reviewed the literature on AI *trust* trajectory. Their review focused primarily on the predicted different trust trajectories of different manifestations of AI; specifically, they examined robotic AI, virtual AI (ie, AI that is a virtual agent or bot, but does not take physical form), and embedded AI (ie, AI that is incorporated in a computer or other tool and therefore invisible to the decision-subject). Some of this theorizing is applicable to the issue of AI fairness perceptions as well. For example, for robotic AI, there tends to be low initial trust that increases over time, whereas for virtual and embedded AI, there tends to be high initial trust that decreases over time. This may extend to fairness perceptions, where robotic AI fairness increases over time and virtual and embedded AI fairness decreases over time.

It is important to note, however, that there is also research that contradicts these general trends (eg, Alan et al., 2014), suggesting that research on AI trust trajectory is in a rudimentary stage and warrants further investigation. One potentially critical factor to explain the trajectory in perceptions of AI is the level of machine intelligence, that is, its technological abilities (higher machine intelligence means the AI is more capable of complex and autonomous actions; Hancock et al., 2011). Glikson and Woolley (2020) suggest that the level of machine intelligence plays a moderating role in trust trajectory, such that trust in AI is more likely to decrease over time when machine intelligence is low. This may be especially true when there is a mismatch between the appearance of the AI system and machine intelligence, such as robots that are more human-like in appearance (and therefore signal intelligence) but are lacking in machine intelligence (Ben Mimoun et al., 2017). Applying this logic to fairness perceptions, it suggests the simple idea that if AI cannot produce the level of machine intelligence that people perceive to be required to make fair judgments and decisions, it will be rejected; for example, if AI is perceived as incapable of discerning a fair (versus unfair) allocation of a task, fairness perceptions may decrease by default. Even if its visual appearance initially signals intelligence, if there is a mismatch in initially perceived and actual machine intelligence, fairness perceptions may suffer over time.

Having said that, future work could look to organizational justice research to provide some direction as to how perceptions of AI procedural and distributive fairness may evolve over time. With time, decision-subjects are likely to acquire more information about how procedures and distributions have been allocated (Ambrose & Cropanzano, 2003). This may involve information acquired about how allocations have been distributed relative to another similarly situated individual or group. In other words, information plays a key role in how perceptions of procedural and distributive fairness evolve over time. Additionally, these perceptions of procedural and distributive fairness may likely depend on the time point at which these perceptions are measured. Perceptions are likely to vary depending upon whether perceptions are measured prior to, immediately after, or well after the allocation decision (Ambrose & Cropanzano, 2003).

5.3. Limitations & reflections

As with any review, the present paper is not without its limitations. Most notably, even though we were able to explicate general trends in the perceived fairness of AI systems through our rigorous, qualitative, narrative review process, we acknowledge that the current state of the literature is not mature enough to quantitatively compare effect sizes across studies, address targeted and highly-specific review questions, and/or derive the kinds of robustly generalizable insights about perceived AI fairness that might be generated during a more systematic review process. Specifically, the current state of the literature appears to be too heterogenous (spanning a variety of theoretical and empirical approaches that are not always directly comparable), and scant (relative to the corpus sizes typically needed for successful systematic reviews) at the moment. However, with enough sustained and rigorous research attention to this topic, systematic and quantitative reviews of the AI fairness literature may become appropriate in due time, and we encourage scholars to pursue such endeavors in the future.

Moreover, given the heterogenous nature of our corpus, we also had to account for variations in the ways in which key concepts were discussed and operationalized across diverse theoretical and empirical traditions. Notably, although our review was focused on the fairness perceptions of "decision-subjects" – ie, those about whom decisions are made – in a few experimental set-ups, the participants whose fairness perceptions were solicited were not directly decision-subjects themselves. Rather, in some studies we examined, participants were asked to evaluate the fairness of AI-based decision-making more generally, or in some simulated context where they were positioned as possible future (rather than current) end-users or decision-subjects of AI-based decision-making systems. Another notable variation pertained to how different studies conceptualized AI involvement in decision-making differently. Some studies were set up as direct comparisons between AI-led and human-led decision-making processes, whereas others included more nuanced conditions (eg, human-led process with minimal AI involvement; AI-led process with a human in the loop, etc.). Although we documented all these differences during our review process, we found no consistent general trends in terms of how fairness perceptions differed according to the level of AI involvement in decision-making, or whose fairness perceptions were being measured (eg, current vs. potential decision-subject; end-user vs. decision-subject, etc.). For these reasons, we did not substantively discuss these dimensions in the main body of our review, although we retain the intuition that these are important dimensions along which fairness perceptions may vary. As such, we encourage future researchers to pay close attention to *whose* perceptions are being measured[7] and *how* exactly AI is being brought into decision-making processes, so as to push forward research on this topic.

6. Conclusion

As organizations increasingly deploy AI systems in critical decision-making functions, those who are affected by these decisions are increasingly concerned about the fairness of these AI systems. As both the scale and scope of AI-augmented decision-making increases, these concerns become ever more pressing. It is therefore important to understand in greater detail how decision-subjects *perceive* the fairness of using AI in decision-making. To address this question, we connected the nascent literature on perceived AI fairness to the well-established and rich organizational literature on perceived fairness. Specifically, we used an established theoretical framework on fairness perceptions to categorize, explicate, and synthesize existing empirical research on perceived AI fairness, and, based on established research findings in the organizational fairness literature, we traced an agenda for future research on perceived AI fairness. In so doing, we hope to stimulate further research on this topic – with interdisciplinary collaborations across organizational studies and human-computer interaction – and in turn, to prompt organizations to pay more careful attention to subjective perceptions of fairness when deploying AI systems in decision-making contexts.

Notes

1. In what follows, we use the term "decision-subject" to refer to those who are affected by AI-augmented decision-making, both (a) *within* organizations (eg, employees), and (b) *outside* organizations (eg, existing or potential customers). During our literature review process, we attempted to distinguish between the two groups, but found no noteworthy differences in their fairness perceptions of AI-augmented decision-making, and as such, we did not press this distinction in our write-up. Our thanks to an anonymous reviewer for pointing out the need for this clarification.
2. Importantly, organizational justice scholars have been relatively slow to attend to issues of perceived AI fairness in their research, despite being especially well-placed to do so (cf. De Cremer, 2020; Robert et al., 2020). As such, by demonstrating the close interconnections between research in organizational justice and AI fairness, we hope that our paper will inspire organizational scholars to better engage with existing scholarship on this topic – especially in the field of human-computer interaction – and in turn, to employ their disciplinary expertise in further advancing scholarship on perceived AI fairness.
3. Integrative reviews, in general, aim at "[reviewing], [critiquing] and [synthesizing] representative literature on a topic in an integrated way such that new frameworks and perspectives on the topic are generated" (Torraco, 2005). Such reviews are often useful for analysing new, emerging topics – to generate initial and/or preliminary conceptualizations, and to combine perspectives from different research traditions (Snyder, 2019). As such, this method appears especially well-suited for reviewing the emerging interdisciplinary topic of perceived AI fairness.
4. We use the term 'organizational justice' to refer to a specific sub-field of organizational scholarship focused on understanding how justice and fairness are perceived by organizational actors. This field has thus far largely treated 'justice' and 'fairness' as equivalent concepts (cf. Colquitt & Rodell, 2015; Greenberg, 2002). However, recent scholarship has argued that, while related, justice and fairness are distinct (Costanza-Chock, 2020; Green, 2022; Kasirzadeh, 2022; Le Bui & Noble, 2020). For the sake of conceptual clarity, throughout this paper, we employ the term '[organizational] justice' to refer to the specific field of scholarship and its associated concepts, and use the term 'fairness' in all other cases.
5. Another important reason why, in our view, Colquitt's framework is well-suited to the purposes of our review pertains to the *interdisciplinary* nature of organizational justice scholarship. Empirical research on how decision-subjects perceive the fairness of AI systems is conducted across a variety of disciplinary traditions with their unique perspectives and theoretical assumptions – psychology, HCI, organizational studies, etc. – and this diversity of perspectives is, in our view, an important reason why the field appears to be currently thriving. As such, it seems unproductive to organize our literature review in ways that reify these disciplinary boundaries: by, for example, sorting our papers according to field, or by using a framework that unduly privileges contributions from one field over all others. Colquitt's framework, given its general focus on categorizing the different types of fairness perceptions from the point of view of the *decision-subject*, retains a broad applicability, and therefore seems especially appropriate for the purposes for our review. Our thanks to an anonymous reviewer for raising this important point.
6. This recently-published paper was not part of our original review corpus, but was instead added later during the peer review process. We thank the anonymous reviewer who brought this paper to our attention.
7. As other recent reviewers of this literature have also observed (cf. van Berkel et al., 2023), it is surprising to note that several extant studies do not report, in sufficient detail, the exact demographic make-up and positionality of their participants. This lack of detail makes it difficult for researchers to fully evaluate the state of current literature, and to identify critical gaps for future research. As such, we echo van Berkel et al.'s call for researchers working on this topic to carefully "report details on recruitment strategy, including compensation and recruitment source (eg, students, crowdworkers), as well as demographic factors of the participant sample (eg, location, age distribution)" in their studies (p. 11).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The first and fourth authors of this paper were supported by a research project grant from the National University of Singapore (NUS) Centre for Trusted Internet and Community (Grant Number: CTIC-RP-20-06) awarded to the last author.

ORCID

Devesh Narayanan http://orcid.org/0000-0003-4201-1421
Mahak Nagpal http://orcid.org/0000-0002-2642-0156
Jack McGuire http://orcid.org/0000-0002-4365-3525
Shane Schweitzer http://orcid.org/0000-0002-4548-0410
David De Cremer http://orcid.org/0000-0002-6357-9385

References

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, *28*(4), 399–416. https://doi.org/10.1111/ijsa.12306

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*. *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267–299). Academic Press.

Ahnert, G., Smirnov, I., Lemmerich, F., Wagner, C., & Strohmaier, M. (2021, June). The FairCeptron: A framework for measuring human perceptions of algorithmic fairness. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 401–403). ACM. https://doi.org/10.1145/3450614.3463291

Alan, A., Costanza, E., Fischer, J., Ramchurn, S. D., Rodden, T., Jennings, N. R. (2014). A field study of human-agent interaction for electricity tariff switching. *13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014)* (pp. 965–972).

Alarie, B., Niblett, A., & Yoon, A. H. (2018). How artificial intelligence will affect the practice of law. *University of Toronto Law Journal*, *68*(Suppl. 1), 106–124. https://doi.org/10.3138/utlj.2017-0052

Albach, M., & Wright, J. R. (2021). The role of accuracy in algorithmic process fairness across multiple domains. In Proceedings of the 22nd ACM Conference on Economics and Computation (pp. 29–49). ACM. https://doi.org/10.1145/3465456.3467620

Ambrose, M. L., & Cropanzano, R. (2003). A longitudinal analysis of organizational fairness: An examination of reactions to tenure and promotion decisions. *The Journal of Applied Psychology*, *88*(2), 266–275. https://doi.org/10.1037/0021-9010.88.2.266

Antifakos, S., Kern, N., Schiele, B., & Schwaninger, A. (2005). Towards improving trust in context-aware systems by displaying system confidence. In Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services, MobileHCI '05 (pp. 9–14). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/1085777.1085780

Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*(C), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bai, B., Dai, H., Zhang, D., Zhang, F., Hu, H. (2021). The impacts of algorithmic work assignment on fairness perceptions and productivity. In *Academy of Management Proceedings Vol. 2021, No. 1*, (pp. 12335). Academy of Management. https://doi.org/10.5465/AMBPP.2021.175

Bankins, S., Formosa, P., Griep, Y., & Richards, D. (2022). AI decision making with dignity? contrasting workers' justice perceptions of human and AI decision making in a human resource management context. *Information Systems Frontiers*, *24*(3), 857–875. https://doi.org/10.1007/S10796-021-10223-8/FIGURES/3

Banks, J. (2021). Good robots, bad robots: Morally valenced behavior effects on perceived mind, morality, and trust. *International Journal of Social Robotics*, *13*(8), 2021–2038. https://doi.org/10.1007/s12369-020-00692-3

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–16). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445717

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the economics of finance* (Vol. 1, pp. 1053–1128). Elsevier.

Barlas, P., Kleanthous, S., Kyriakou, K., & Otterbacher, J. (2019, June). What makes an image tagger fair? In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 95–103). ACM. https://doi.org/10.1145/3320435.3320442

Barnett, A., Savic, M., Pienaar, K., Carter, A., Warren, N., Sandral, E., Manning, V., & Lubman, D. I. (2021). Enacting 'more-than-human'-care: Clients' and counsellors' views on the multiple affordances of chatbots in alcohol and other drug counselling. *The International Journal on Drug Policy*, *94*(3), 102910. https://doi.org/10.1016/j.drugpo.2020.102910

Barsky, A., & Kaplan, S. A. (2007). If you feel bad, it's unfair: A quantitative synthesis of affect and organizational justice perceptions. *The Journal of Applied Psychology*, *92*(1), 286–295. https://doi.org/10.1037/0021-9010.92.1.286

Ben Mimoun, M. S., Poncin, I., & Garnier, M. (2017). Animated conversational agents and e- consumer productivity: The roles of agents and individual characteristics. *Information and Management*, *54*(5), 545–559. https://doi.org/10.1016/j.im.2016.11.008

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial intelligence in organizations: Current state and future opportunities. *MIS Quarterly Executive*, *19*(4), 9–21. https://doi.org/10.2139/ssrn.3741983

Bies, R. J., & Moag, J. F. (1986). Interactional justice: Communication criteria of fairness. In R. J. Lewicki, B. H. Sheppard, & M. H. Bazerman (Eds.), *Research on negotiations in organizations* (Vol. 1, pp. 43–55). JAI Press.

Binns, R. (2020, January). On the apparent conflict between individual and group fairness. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 514–524). ACM. https://doi.org/10.1145/3351095.3372864

Binns, R. (2022a). Analogies and disanalogies between machine-driven and human-driven legal judgement. *Journal of Cross-Disciplinary Research in Computational Law*, *1*(1). https://journalcrcl.org/crcl/article/view/5

Binns, R. (2022b). Human Judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance*, *16*(1), 197–211. https://doi.org/10.1111/rego.12358

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., Shadbolt, N. (2018, April). 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* (pp. 1–14). ACM.

Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., Walker, K. (2020). *Fairlearn: A toolkit for assessing and improving fairness in AI*. [Microsoft, Technical Report No. MSR-TR-2020-32].

Brockner, J., De Cremer, D., van Dijke, M., De Schutter, L., Holtz, B., & Van Hiel, A. (2021). Factors affecting supervisors' enactment of interpersonal fairness: The interactive relationship between their managers' informational fairness and supervisors' sense of power. *Journal of Organizational Behavior*, *42*(6), 800–813. https://doi.org/10.1002/job.2466

Brockner, J., Konovsky, M., Cooper-Schneider, R., Folger, R., Martin, C., & Bies, R. J. (1994). Interactive effects of procedural justice and outcome negativity on victims and survivors of job loss. *Academy of Management Journal*, *37*(2), 397–409. https://doi.org/10.2307/256835

Brown, A., Chouldechova, A., Putnam-Hornstein, E., Tobin, A., Vaithianathan, R. (2019, May). Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). ACM.

Butcher, J., & Beridze, I. (2019). What is the state of artificial intelligence governance globally? *The RUSI Journal*, *164*(5–6), 88–96. https://doi.org/10.1080/03071847.2019.1694260

Camerer, C. (1999). Behavioral economics: Reunifying psychology and economics. Proceedings of the National Academy of Sciences, 96(19), 10575–10577. https://doi.org/10.1073/pnas.96.19.10575

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, *538*(7623), 20–23. https://doi.org/10.1038/538020a

Chang, M. L., Trafton, G., McCurry, J. M., & Thomaz, A. L. (2021, August). Unfair! perceptions of fairness in human-robot teams. In 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN (pp. 905–912). IEEE. https://doi.org/10.1109/RO-MAN50785.2021.9515428

Charisi, V., Imai, T., Rinta, T., Nakhayenze, J. M., & Gomez, R. (2021, June). *Exploring the concept of fairness in everyday, imaginary and robot scenarios: A cross-cultural study with children in Japan and Uganda* [Paper presentation]. Interaction Design and Children (pp. 532–536). https://doi.org/10.1145/3459990.3465184

Chen, B. M., Stremitzer, A., & Tobia, K. (2021). Having your day in robot court. UCLA School of Law, Public Law Research Paper (pp. 21–20).

Cheng, H.-F., Stapleton, L., Wang, R., Bullock, P., Chouldechova, A., Wu, Z. S. S., & Zhu, H. (2021). Soliciting stakeholders' fairness notions in child maltreatment predictive systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–17). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445308

Choi, S., Mattila, A. S., & Bolton, L. E. (2021). To err is human (-oid): How do consumers react to robot service failure and recovery? *Journal of Service Research*, *24*(3), 354–371. https://doi.org/10.1177/1094670520978798

Chouldechova, A., Benavides-Prado, D., Fialko, O., & Vaithianathan, R. (2018). *A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions* [Paper presentation]. Conference on Fairness, Accountability and Transparency (pp. 134–148). PMLR.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *The Journal of Applied Psychology*, *86*(3), 386–400. https://doi.org/10.1037/0021-9010.86.3.386

Colquitt, J. A., & Rodell, J. B. (2015). Chapter 8. Measuring justice and fairness. In R. S. Cropanzano & M. L. Ambrose (eds). *The Oxford Handbook of Justice in the Workplace* (pp. 187). Oxford University Press.

Colquitt, J. A., Hill, E. T., & De Cremer, D. (2023). Forever focused on fairness: 75 years of organizational justice in personnel psychology. *Personnel Psychology*, *76*(2), 413–435. https://doi.org/10.1111/peps.12556

Colquitt, J. A., Scott, B. A., Rodell, J. B., Long, D. M., Zapata, C. P., Conlon, D. E., & Wesson, M. J. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *The Journal of Applied Psychology*, *98*(2), 199–236. https://doi.org/10.1037/a0031757

Cook, K. S., & Hegtvedt, K. A. (1983). Distributive justice, equity, and equality. *Annual Review of Sociology*, *9*(1), 217–241. https://doi.org/10.1146/annurev.so.09.080183.001245

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Part F1296 (pp. 797–806). ACM. https://doi.org/10.1145/3097983.3098095

Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. MIT Press.

Cowgill, B. (2018). *Bias and productivity in humans and algorithms: Theory and evidence from resume screening* (pp. 29). Columbia Business School, Columbia University.

Cropanzano, R., Rupp, D. E., Mohler, C. J., & Schminke, M. (2001). Three roads to organizational justice. In G. R. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 20, pp. 1–123). Elsevier Science/JAI Press. https://doi.org/10.1016/S0742-7301(01)20001-2

Daileyl, R. C., & Kirk, D. J. (1992). Distributive and procedural justice as antecedents of job dissatisfaction and intent to turnover. *Human Relations*, *45*(3), 305–317. https://doi.org/10.1177/001872679204500306

Daly, J. P., & Tripp, T. M. (1996). Is outcome fairness used to make procedural fairness judgments when procedural information is inaccessible? *Social Justice Research*, *9*(4), 327–349. https://doi.org/10.1007/BF02196989

De Cremer, D. (2007). Emotional effects of distributive justice as a function of autocratic leader behavior. *Journal of Applied Social Psychology*, *37*(6), 1385–1404. https://doi.org/10.1111/j.1559-1816.2007.00217.x

De Cremer, D. (2020). What does building a fair AI really entail? *Harvard Business Review*

De Cremer, D., & Chun, J. (2021). *Algorithmic evaluation and its unfairness: The centrality of respect and the lack thereof* [Unpublished manuscript].

De Cremer, D., & De Schutter, L. (2021). How to use algorithmic decision-making to promote inclusiveness in organizations. *AI and Ethics*, *1*(4), 563–567. https://doi.org/10.1007/s43681-021-00073-0

De Cremer, D., Kasparov, G. (2021). AI should augment human intelligence, not replace it. *Harvard Business Review*.

De Cremer, D., & McGuire, J. (2022). Human–algorithm collaboration works best if humans lead (because it is fair!). *Social Justice Research*, *35*(1), 33–55. https://doi.org/10.1007/s11211-021-00382-z

De Cremer, D., & Tyler, T. R. (2005). Managing group behavior: The interplay between procedural justice, sense of self, and cooperation. *Advances in Experimental Social Psychology*, *37*, 151–218. https://doi.org/10.1016/S0065-2601(05)37003-1

De Cremer, D., Brockner, J., Fishman, A., van Dijke, M., van Olffen, W., & Mayer, D. M. (2010). When do procedural fairness and outcome fairness interact to influence employees' work attitudes and behaviors? The moderating effect of uncertainty. *The Journal of Applied Psychology*, *95*(2), 291–304. https://doi.org/10.1037/a0017866

De Cremer, D., Narayanan, D., Deppeler, A., Nagpal, M., & McGuire, J. (2022). The road to a human-centred digital society: Opportunities, challenges and responsibilities for humans in the age of machines. *AI and Ethics*, *2*(4), 579–583. https://doi.org/10.1007/s43681-021-00116-6

De Cremer, D., Zeelenberg, M., & Murnighan, J. K. (2013). *Social Psychology and Economics*. Psychology Press.

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. *AI & Society*, *35*(4), 917–926. https://doi.org/10.1007/s00146-020-00960-w

Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, *31*(3), 137–149. https://doi.org/10.1111/j.1540-4560.1975.tb01000.x

DeVito, M. A., Gergle, D., & Birnholtz, J. (2017). Algorithms ruin everything': #RIPTwitter, folk theories, and resistance to algorithmic change in social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* CHI '17. (pp. 3163–3174). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3025453.3025659

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, *5*(7), 809–828. https://doi.org/10.1080/21670811.2016.1208053

Dietvorst, B. J., & Bartels, D. M. (2022). Consumers object to algorithms making morally relevant tradeoffs because of algorithms' consequentialist decision strategies. *Journal of Consumer Psychology*, *32*(3), 406–424. https://doi.org/10.1002/jcpy.1266

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science*, *31*(10), 1302–1314. https://doi.org/10.1177/0956797620948841

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Dietz, J., Robinson, S. L., Folger, R., Baron, R. A., & Schulz, M. (2003). The impact of community violence and an organization's procedural justice climate on workplace aggression. *Academy of Management Journal*, *46*(3), 317–326. https://doi.org/10.2307/30040625

Dineen, B. R., Noe, R. A., & Wang, C. (2004). Perceived fairness of web-based applicant screening procedures: Weighing the rules of justice and the role of individual differences. *Human Resource Management*, *43*(2–3), 127–145. https://doi.org/10.1002/hrm.20011

Dodge, J., Liao, Q. V., Zhang, Y., Bellamy, R. K., Dugan, C. (2019, March). Explaining models: An empirical study of how explanations impact fairness judgment. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 275–285). ACM.

Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke Law & Technology Review's*, *16*, 18–84. https://doi.org/10.31228/osf.io/97upg

Elahi, M., Abdollahpouri, H., Mansoury, M., & Torkamaan, H. (2021, June). Beyond algorithmic fairness in recommender systems. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 41–46). ACM. https://doi.org/10.1145/3450614.3461685

Eslami, M., Vaccaro, K., Lee, M. K., Elazari Bar On, A., Gilbert, E., & Karahalios, K. (2019). User attitudes towards algorithmic opacity and transparency in online reviewing platforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3290605.3300724

Ferraro, A., Serra, X., & Bauer, C. (2021, August). *What is fair? Exploring the artists' perspective on the fairness of music streaming platforms*. In *IFIP Conference on Human-Computer Interaction* (pp. 562–584). Springer.

Firestone, C. (2020). Performance vs. competence in human–machine comparisons. Proceedings of the National Academy of Sciences, 117(43), 26562–26571. https://doi.org/10.1073/pnas.1905334117

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center Research Publication.

Fleisher, W. (2021, July). What's fair about individual fairness? In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 480–490). ACM. https://doi.org/10.1145/3461702.3462621

Fleiß, J., Bäck, E., Thalmann, S. (2020). Explainability and the intention to use AI-based conversational agents. In *Proceedings of the First International Workshop on Explainable and Interpretable Machine Learning (XI-ML 2020).*

Folger, N., Brosi, P., Stumpf-Wollersheim, J., & Welpe, I. M. (2022). Applicant reactions to digital selection methods: A signaling perspective on innovativeness and procedural justice. *Journal of Business and Psychology*, *37*(4), 735–757. https://doi.org/10.1007/s10869-021-09770-3

Formosa, P., Rogers, W., Griep, Y., Bankins, S., & Richards, D. (2022). Medical AI and human dignity: Contrasting perceptions of human and artificially intelligent (AI) decision making in diagnostic and medical resource allocation contexts. *Computers in Human Behavior*, *133*, 107296. https://doi.org/10.1016/j.chb.2022.107296

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organization technology isn't the biggest challenge, culture is. *Harvard Business Review*, *97*(4), 62–74. https://hbr.org/2019/07/building-the-ai-powered-organization

Gamez, P., Shank, D. B., Arnold, C., & North, M. (2020). Artificial virtue: The machine question and perceptions of moral character in artificial moral agents. *AI & Society*, *35*(4), 795–809. https://doi.org/10.1007/s00146-020-00977-1

Gino, F., & Pisano, G. (2008). Toward a theory of behavioral operations. *Manufacturing & Service Operations Management*, *10*(4), 676–691. https://doi.org/10.1287/msom.1070.0205

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goldfarb, A., & Lindsay, J. (2020). *Artificial intelligence in war: Human judgment as an organizational strength and a strategic liability*. Brookings Institution.

Goldman, E. (2005). Search engine bias and the demise of search engine utopianism. *Yale JL & Tech*, *8*, 188. http://hdl.handle.net/20.500.13051/7858

Gonçalves, J., Weber, I., Masullo, G. M., Torres da Silva, M., & Hofhuis, J. (2021). Common sense or censorship: How algorithmic moderators and message type influence perceptions of online content deletion. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448211032310

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Green, B. (2020, January). The false promise of risk assessments: Epistemic reform and the limits of fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 594–606). ACM.

Green, B. (2022). Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology*, *35*(4), 90. https://doi.org/10.1007/s13347-022-00584-6

Green, B., & Chen, Y. (2019). Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In FAT* 2019 – Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (pp. 90–99). ACM. https://doi.org/10.1145/3287560.3287563

Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management*, *16*(2), 399–432. https://doi.org/10.1177/014920639001600208

Greenberg, J. (1993). Stealing in the name of justice: Informational and interpersonal moderators of theft reactions to underpayment inequity. *Organizational Behavior and Human Decision Processes*, *54*(1), 81–103. https://doi.org/10.1006/obhd.1993.1004

Greenberg, J. (2002). *Advances in organizational justice*. Stanford University Press.

Greenberg, J., & Cropanzano, R. (1993). The social side of fairness: Interpersonal and informational classes of organizational justice. *Justice in the workplace: Approaching fairness in human resource management*. Lawrence Erlbaum Associates.

Grgić-Hlača, N., Redmiles, E. M., Gummadi, K. P., & Weller, A. (2018a). Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In *Proceedings of the 2018 World Wide Web Conference* (pp. 903–912). ACM.https://doi.org/10.1145/3178876.3186138

Grgić-Hlača, N., Weller, A., & Redmiles, E. M. (2020). Dimensions of diversity in human perceptions of algorithmic fairness. *arXiv preprint arXiv:2005.00808.*

Grgić-Hlača, N., Zafar, M. B., Gummadi, K. P., Weller, A. (2018b, April). Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1). https://doi.org/10.1609/aaai.v32i1.11296

Grudin, J. (1996). The organizational contexts of development and use. *ACM Computing Surveys*, *28*(1), 169–171. https://doi.org/10.1145/234313.234384

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, *4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Gupta, M., Parra, C. M., & Dennehy, D. (2021). Questioning racial and gender bias in AI-based recommendations: Do espoused national cultural values matter? *Information Systems Frontiers*, *24*(5), 1465–1481. https://doi.org/10.1007/s10796-021-10156-2

Hakami, E., & Hernández Leo, D. (2020). How are learning analytics considering the societal values of fairness, accountability, transparency and human well-being?: A literature review. In A. Martínez-Monés, A. Álvarez, M. Caeiro-Rodríguez, Y. Dimitriadis, (Eds.). *LASI-SPAIN 2020: Learning analytics summer institute Spain 2020: Learning analytics. Time for adoption?* (p. 121–141). CEUR.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human–robot interaction. *Human Factors*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Hannan, J., Chen, H. Y. W., & Joseph, K. (2021, July). Who gets what, according to whom? an analysis of fairness perceptions in service allocation. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 555–565). ACM.https://doi.org/10.1145/3461702.3462568

Harrison, G., Hanson, J., Jacinto, C., Ramirez, J., & Ur, B. (2020). An empirical study on the perceived fairness of realistic, imperfect machine learning models. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 392–402). ACM. https://doi.org/10.1145/3351095.3372831

Hase, P., & Bansal, M. (2020). Evaluating explainable AI: Which algorithmic explanations help users predict model behavior? In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5540–5552). https://doi.org/10.18653/v1/2020.acl-main.491

Helberger, N., Araujo, T., & de Vreese, C. H. (2020). Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. *Computer Law & Security Review*, *39*, 105456. https://doi.org/10.1016/j.clsr.2020.105456

Hobson, Z., Yesberg, J. A., Bradford, B., & Jackson, J. (2021). Artificial fairness? Trust in algorithmic police decision-making. *Journal of Experimental Criminology*, *19*(1), 165–189.

Höddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, *116*, 106635. https://doi.org/10.1016/j.chb.2020.106635

Holstein, K., Wortman Vaughan, J., Daumé, H., III, Dudik, M., Wallach, H. (2019, May). Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). ACM.

Htun, N. N., Lecluse, E., & Verbert, K. (2021, April). Perception of fairness in group music recommender systems. In *26th International Conference on Intelligent User Interfaces* (pp. 302–306). ACM. https://doi.org/10.1145/3397481.3450642

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics*, *178*(4), 977–1007. https://doi.org/10.1007/s10551-022-05049-6

Jarrahi, M. H., Newlands, G., Lee, M. K., Wolf, C. T., Kinder, E., & Sutherland, W. (2021). Algorithmic management in a work context. *Big Data & Society*, *8*(2), 205395172110203. 20539517211020332. https://doi.org/10.1177/20539517211020332

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Jones, D. A., & Skarlicki, D. P. (2013). How perceptions of fairness can change: A dynamic model of organizational justice. *Organizational Psychology Review*, *3*(2), 138–160. https://doi.org/10.1177/2041386612461665

Kaibel, C., Koch-Bayram, I., Biemann, T., Mühlenbock, M. (2019). Applicant perceptions of hiring algorithms-uniqueness and discrimination experiences as moderators. Academy of Management Annual Meeting Proceedings, 2019(1), 18172. https://doi.org/10.5465/AMBPP.2019.210

Karizat, N., Delmonaco, D., Eslami, M., Andalibi, N. (2021). Algorithmic folk theories and identity: How TikTok users co-produce knowledge of identity and engage in algorithmic resistance. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1–44. https://doi.org/10.1145/3476046

Kasinidou, M., Kleanthous, S., Barlas, P., & Otterbacher, J. (2021). I agree with the decision, but they didn't deserve this: future developers' perception of fairness in algorithmic decisions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 690–700). Canada: ACM. https://doi.org/10.1145/3442188.3445931

Kasirzadeh, A. (2022). Algorithmic fairness and structural injustice: Insights from feminist political philosophy. *arXiv preprint arXiv: 2206.00945*. https://doi.org/10.1145/3514094.3534188

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., Wortman Vaughan, J. (2020). Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). ACM.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Kieslich, K., Keller, B., & Starke, C. (2022). Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence. *Big Data & Society*, *9*(1), 205395172210929. https://doi.org/10.1177/20539517221092956

Kim, T. Y., & Leung, K. (2007). Forming and reacting to overall fairness: A cross-cultural comparison. *Organizational Behavior and Human Decision Processes*, *104*(1), 83–95. https://doi.org/10.1016/j.obhdp.2007.01.004

Kleanthous, S., Kasinidou, M., Barlas, P., & Otterbacher, J. (2022). Perception of fairness in algorithmic decisions: Future developers' perspective. *Patterns*, *3*(1), 100380. https://doi.org/10.1016/j.patter.2021.100380

Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, *114*, 106553. https://doi.org/10.1016/j.chb.2020.106553

Konradt, U., Garbers, Y., Erdogan, B., & Bauer, T. (2016). Patterns of change in fairness perceptions during the hiring process. *International Journal of Selection and Assessment*, *24*(3), 246–259. https://doi.org/10.1111/ijsa.12144

Kushwaha, A. K., Pharswan, R., & Kar, A. K. (2021). Always trust the advice of AI in difficulties? Perceptions around AI in decision making. In *Conference on e-Business, e-Services and e-Society* (pp. 132–143). Springer.

Kuutti, K., Bannon, L. J. (2014, April). The turn to practice in HCI: Towards a research agenda. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3543–3552). ACM.

Lai, M. C., Brian, M., & Mamzer, M. F. (2020). Perceptions of artificial intelligence in healthcare: Findings from a qualitative survey study among actors in France. *Journal of Translational Medicine*, *18*(1), 1–13. https://doi.org/10.1186/s12967-019-02204-y

Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*, *123*(4), 106878. https://doi.org/10.1016/j.chb.2021.106878

Langer, M., Baum, K., König, C. J., Hähne, V., Oster, D., & Speith, T. (2021). Spare me the details: How the type of information about automated interviews influences applicant reactions. *International Journal of Selection and Assessment*, *29*(2), 154–169. https://doi.org/10.1111/ijsa.12325

Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, *27*(3), 217–234. https://doi.org/10.1111/ijsa.12246

Langer, M., König, C. J., Back, C., & Hemsing, V. (2022). Trust in Artificial Intelligence: Comparing trust processes between human and automated trustees in light of unfair bias. *Journal of Business and Psychology*, *38*(2), 1–16. https://doi.org/10.1007/s10869-022-09829-9

Latonero, M. (2018). Governing artificial intelligence: Upholding human rights & dignity. *Data & Society*, 1–37. https://datasociety.net/library/governing-artificial-intelligence/

Le Bui, M., & Noble, S. U. (2020). We're missing a moral framework of justice in artificial intelligence. In Dubber, M. D., Pasquale, F., & Das, S. (Eds.), *The Oxford handbook of ethics of AI*, (pp. 163–179). Oxford University Press.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 205395171875668. 2053951718756684. https://doi.org/10.1177/2053951718756684

Lee, M. K., Baykal, S. (2017, February). Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1035–1048). ACM.

Lee, M. K., & Rich, K. (2021, May). *Who is included in human perceptions of AI?: Trust and perceived fairness around healthcare AI and cultural mistrust. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). https://doi.org/10.1145/3411764.3445570

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., Kusbit, D. (2019a). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–26. https://doi.org/10.1145/3359284

Lee, M. K., Kusbit, D., Kahng, A., Kim, J. T., Yuan, X., Chan, A., See, D. (2019b). WeBuildAI: Participatory Framework for Algorithmic Governance. Proceedings of the ACM on Human-Computer Interaction, *3*(CSCW), 1–35. https://doi.org/10.1145/3359283

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, *31*(4), 611–627. https://doi.org/10.1007/s13347-017-0279-x

Leventhal, G. S. (1976). The distribution of rewards and resources in groups and organizations. In L. Berkowitz & W. Walster (Eds.), *Advances in experimental social psychology* (Vol. 9, pp. 91–131). Academic Press.

Li, L., Lassiter, T., Oh, J., Lee, M. K. (2021, July). Algorithmic hiring in practice: Recruiter and HR professional's perspectives on AI use in hiring. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 166–176). ACM.

Lima, G., Grgić-Hlača, N., & Cha, M. (2021, May). *Human perceptions on moral responsibility of AI: A case study in AI-assisted bail decision-making*. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–17). ACM. https://doi.org/10.1145/3411764.3445260

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. Springer Science & Business Media.

Litoiu, A., Ullman, D., Kim, J., & Scassellati, B. (2015, March). *Evidence that robots trigger a cheating detector in humans*. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 165–172). https://doi.org/10.1145/2696454.2696456

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

Lyons, H., Velloso, E., & Miller, T. (2021). Fair and Responsible AI: A focus on the ability to contest. *arXiv preprint arXiv:2102.10787*

Marcinkowski, F., Kieslich, K., Starke, C., Lünich, M. (2020, January). 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* (pp. 122–130). ACM.

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, *160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

Masrour, F., Tan, P. N., & Esfahanian, A. H. (2020, November). Fairness perception from a network-centric perspective. In *2020 IEEE International Conference on Data Mining (ICDM)* (pp. 1178–1183). IEEE. https://doi.org/10.1109/ICDM50108.2020.00145

McGuire, J., & De Cremer, D. (2022). Algorithms, leadership, and morality: Why a mere human effect drives the preference for human over algorithmic leadership. *AI and Ethics*. Advance online publication. https://doi.org/10.1007/s43681-022-00192-2

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607

Miller, S. M., & Keiser, L. R. (2021). Representative bureaucracy and attitudes toward automated decision making. *Journal of Public Administration Research and Theory*, *31*(1), 150–165. https://doi.org/10.1093/jopart/muaa019

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*(C), 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Mirowska, A., & Mesnet, L. (2022). Preferring the devil you know: Potential applicant reactions to artificial intelligence evaluation of interviews. *Human Resource Management Journal*, *32*(2), 364–383. https://doi.org/10.1111/1748-8583.12393

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, *8*(1), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 205395171667967. https://doi.org/10.1177/2053951716679679

Morse, L., Teodorescu, M. H. M., Awwad, Y., & Kane, G. C. (2022). Do the ends justify the means? Variation in the distributive and procedural fairness of machine learning algorithms. *Journal of Business Ethics*, *181*(4), 1083–1095. https://doi.org/10.1007/s10551-021-04939-5

Nagtegaal, R. (2021). The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly*, *38*(1), 101536. https://doi.org/10.1016/j.giq.2020.101536

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, *160*, 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008

Noble, S. M., Foster, L. L., & Craig, S. B. (2021). The procedural and interpersonal justice of automated application and resume screening. *International Journal of Selection and Assessment*, *29*(2), 139–153. https://doi.org/10.1111/ijsa.12320

Noble, S. U. (2013). Google search: Hyper-visibility as a means of rendering black women and girls invisible. *InVisible Culture*, *19*. https://doi.org/10.47761/494a02f6.50883fff

Nørskov, S., Damholdt, M. F., Ulhøi, J. P., Jensen, M. B., Ess, C., & Seibt, J. (2020). Applicant fairness perceptions of a robot-mediated job interview: A video vignette-based experimental survey. *Frontiers in Robotics and AI*, *7*(163), 586263. https://doi.org/10.3389/frobt.2020.586263

Ogunniye, G., Legastelois, B., Rovatsos, M., Dowthwaite, L., Portillo, V., Perez Vallejos, E., Zhao, J., & Jirotka, M. (2021). Understanding user perceptions of trustworthiness in E-recruitment systems. *IEEE Internet Computing*, *25*(6), 23–32. https://doi.org/10.1109/MIC.2021.3115670

Ötting, S. K., & Maier, G. W. (2018). The importance of procedural justice in human–machine interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior*, *89*(479), 27–39. https://doi.org/10.1016/j.chb.2018.07.022

Park, H., Ahn, D., Hosanagar, K., & Lee, J. (2021, May). Human-AI interaction in human resource management: Understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). ACM. https://doi.org/10.1145/3411764.3445304

Pierson, E. (2017). Demographics and discussion influence views on algorithmic fairness. *arXiv preprint arXiv:1712.09124*

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI*

Conference on Human Factors in Computing Systems (pp. 1–13). ACM. https://doi.org/10.1145/3173574.3173677

Rader, E., & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the facebook news feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI'15* (pp. 173–82). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2702123.2702174

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, *169*(12), 866–872. https://doi.org/10.7326/M18-1990

Renier, L. A., Mast, M. S., & Bekbergenova, A. (2021). To err is human, not algorithmic–Robust reactions to erring algorithms. *Computers in Human Behavior*, *124*(C), 106879. https://doi.org/10.1016/j.chb.2021.106879

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human–Computer Interaction*, *35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391

Rolland, F., & Steiner, D. D. (2007). Test-taker reactions to the selection process: Effects of outcome favorability, explanations, and voice on fairness perceptions. *Journal of Applied Social Psychology*, *37*(12), 2800–2826. https://doi.org/10.1111/j.1559-1816.2007.00282.x

Saha, D., Schumann, C., Mcelfresh, D., Dickerson, J., Mazurek, M., Tschantz, M. (2020, November). Measuring non-expert comprehension of machine learning fairness metrics. In *International Conference on Machine Learning* (pp. 8377–8387). ACM.

Saxena, N. A., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D. C., & Liu, Y. (2020). How do fairness definitions fare? Testing public attitudes towards three algorithmic definitions of fairness in loan allocations. *Artificial Intelligence*, *283*(5), 103238. https://doi.org/10.1016/j.artint.2020.103238

Schadenberg, B. R., Reidsma, D., Heylen, D. K., & Evers, V. (2021). "I see what you did there" understanding people's social perception of a robot and its predictability. *ACM Transactions on Human-Robot Interaction*, *10*(3), 1–28. https://doi.org/10.1145/3461534

Schick, J., & Fischer, S. (2021). Dear computer on my desk, which candidate fits best? An assessment of candidates' perception of assessment quality when using AI in personnel selection. *Frontiers in Psychology*, *12*, 4868. https://doi.org/10.3389/fpsyg.2021.739711

Schlicker, N., Langer, M., Ötting, S. K., Baum, K., König, C. J., & Wallach, D. (2021). What to expect from opening up 'black boxes'? Comparing perceptions of justice between human and automated agents. *Computers in Human Behavior*, *122*(4), 106837. https://doi.org/10.1016/j.chb.2021.106837

Schoeffer, J., & Kuehl, N. (2021, October). Appropriate fairness perceptions? On the effectiveness of explanations in enabling people to assess the fairness of automated decision systems. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (pp. 153–157). ACM.https://doi.org/10.1145/3462204.3481742

Schoeffer, J., Machowski, Y., & Kuehl, N. (2021). A study on fairness and trust perceptions in automated decision making. *arXiv preprint arXiv:2103.04757*

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 59–68). ACM. https://doi.org/10.1145/3287560.3287598

Shandilya, A., Dash, A., Chakraborty, A., Ghosh, K., & Ghosh, S. (2021). Fairness for whom? Understanding the reader's perception of fairness in text summarization. *arXiv preprint arXiv:2101.12406*

Shapiro, D. L., Buttner, E. H., & Barry, B. (1994). Explanations: What factors enhance their perceived adequacy? *Organizational Behavior and Human Decision Processes*, *58*(3), 346–368. https://doi.org/10.1006/obhd.1994.1041

Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media*, *64*(4), 541–565. https://doi.org/10.1080/08838151.2020.1843357

Shin, D. (2022). How do people judge the credibility of algorithmic sources? *AI & Society*, *37*(1), 81–96. https://doi.org/10.1007/s00146-021-01158-4

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*(1), 277–284. https://doi.org/10.1016/j.chb.2019.04.019

Shin, D., Zaid, B., & Ibahrine, M. (2020, November). Algorithm appreciation: algorithmic performance, developmental processes, and user interactions. In *2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)* (pp. 1–5). IEEE. https://doi.org/10.1109/CCCI49893.2020.9256470

Short, E., Hart, J., Vu, M., & Scassellati, B. (2010, March). No fair!! an interaction with a cheating robot. 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI (pp. 219–226). IEEE. https://doi.org/10.1109/HRI.2010.5453193

Shulner-Tal, A., Kuflik, T., & Kliger, D. (2022). Fairness, explainability and in-between: Understanding the impact of different explanation methods on non-expert users' perceptions of fairness toward an algorithmic system. *Ethics and Information Technology*, *24*(1), 1–13. https://doi.org/10.1007/s10676-022-09623-4

Simshaw, D. (2018). Ethical issues in robo-lawyering: The need for guidance on developing and using artificial intelligence in the practice of law. *Hastings Law Journal*, *70*(1), 173. https://repository.uclawsf.edu/hastings_law_journal/vol70/iss1/4

Skewes, J., Amodio, D. M., & Seibt, J. (2019). Social robotics and the modulation of social perception and bias. *Philosophical Transactions of the Royal Society of London Series B*, *374*(1771), 20180037. https://doi.org/10.1098/rstb.2018.0037

Skinner, Z., Brown, S., & Walsh, G. (2020, April). Children of color's perceptions of fairness in AI: An exploration of equitable and inclusive co-design. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, (pp. 1–8). ACM.https://doi.org/10.1145/3334480.3382901

Smith, J., Sonboli, N., Fiesler, C., & Burke, R. (2020). Exploring user opinions of fairness in recommender systems. *arXiv preprint arXiv: 2003.06461.*

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Sonboli, N., Smith, J. J., Cabral Berenfus, F., Burke, R., & Fiesler, C. (2021, June). *Fairness and transparency in recommendation: The users' perspective*. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 274–279). ACM.https://doi.org/10.1145/3450613.3456835

Srivastava, M., Heidari, H., Krause, A. (2019, July). Mathematical notions vs. human perception of fairness: A descriptive approach to fairness for machine learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2459–2468). ACM.

Stai, B., Heller, N., McSweeney, S., Rickman, J., Blake, P., Vasdev, R., Edgerton, Z., Tejpaul, R., Peterson, M., Rosenberg, J., Kalapara, A., Regmi, S., Papanikolopoulos, N., & Weight, C. (2020). Public perceptions of artificial intelligence and robotics in medicine. *Journal of Endourology*, *34*(10), 1041–1048. https://doi.org/10.1089/end.2020.0137

Stapleton, L., Lee, M. H., Qing, D., Wright, M., Chouldechova, A., Holstein, K., Wu, Z. S., & Zhu, H. (2022). Imagining new futures beyond predictive systems in child welfare: A qualitative study with impacted stakeholders. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1162–1177). ACM. https://doi.org/10.1145/3531146.3533177

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2021). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *arXiv preprint arXiv:2103.12016*

Stellmach, H., & Lindner, F. (2019). Perception of an uncertain ethical reasoning robot. *i-com, 18*(1), 79–91. https://doi.org/10.1515/icom-2019-0002

Streicher, B., Jonas, E., Maier, G. W., Frey, D., & Spießberger, A. (2012). Procedural fairness and creativity: Does voice maintain people's creative vein over time? *Creativity Research Journal*, *24*(4), 358–363. https://doi.org/10.1080/10400419.2012.730334

Suen, H. Y., Chen, M. Y. C., & Lu, S. H. (2019). Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? *Computers in Human Behavior*, *98*, 93–101. https://doi.org/10.1016/j.chb.2019.04.012

Swart, J. (2021). Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. *Social Media + Society*, *7*(2), 205630512110088. https://doi.org/10.1177/20563051211008828

Telkamp, J. B., & Anderson, M. H. (2022). The Implications of diverse human moral foundations for assessing the ethicality of artificial intelligence. *Journal of Business Ethics*, *178*(4), 961–976. https://doi.org/10.1007/s10551-022-05057-6

Thibaut, J., & Walker, L. (1978). A theory of procedure. *California Law Review*, *66*(3), 541. https://doi.org/10.2307/3480099

Tomaino, G., Abdulhalim, H., Kireyev, P., & Wertenbroch, K. (2020). *Denied by an (Unexplainable) Algorithm: Teleological Explanations for Algorithmic Decisions Enhance Customer Satisfaction* (SSRN Scholarly Paper ID 3683754). Social Science Research Network. https://doi.org/10.2139/ssrn.3683754

Törnblom, K. Y., & Vermunt, R. (1999). An integrative perspective on social justice: Distributive and procedural fairness evaluations of positive and negative outcome allocations. *Social Justice Research*, *12*(1), 39–64. https://doi.org/10.1023/A:1023226307252

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, *4*(3), 356–367. https://doi.org/10.1177/1534484305278283

Tulk, S., Wiese, E. (2018). Trust and approachability mediate social decision making in human-robot interaction. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62(1), 704–708. https://doi.org/10.1177/1541931218621160

Tyler, T. R., & Bies, R. J. (2015). Beyond formal procedures: The interpersonal context of procedural justice. In *Applied social psychology and organizational settings* (pp. 77–98). Psychology Press.

Vaccaro, K., Sandvig, C., Karahalios, K. (2020). "At the End of the Day Facebook Does What ItWants" How users experience contesting algorithmic content moderation. *Proceedings of the ACM on Human-Computer Interaction*, *4*, 1–22. https://doi.org/10.1145/3415238

Vaccaro, M., & Waldo, J. (2019). The effects of mixing machine learning and human judgment. *Communications of the ACM*, *62*(11), 104–110. https://doi.org/10.1145/3359338

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, *22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

van Berkel, N., Goncalves, J., Hettiachchi, D., Wijenayake, S., Kelly, R. M., Kostakos, V. (2019). Crowdsourcing perceptions of fair predictors for machine learning: A recidivism case study. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–21. https://doi.org/10.1145/3359130

van Berkel, N., Goncalves, J., Russo, D., Hosio, S., & Skov, M. B. (2021). Effect of information presentation on fairness perceptions of machine learning predictors. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3411764.3445365

van Berkel, N., Sarsenbayeva, Z., & Goncalves, J. (2023). The methodology of studying fairness perceptions in Artificial Intelligence: Contrasting CHI and FAccT. *International Journal of Human-Computer Studies*, *170*(C), 102954. https://doi.org/10.1016/j.ijhcs.2022.102954

van Berkel, N., Tag, B., Goncalves, J., & Hosio, S. (2022). Human-centred artificial intelligence: A contextual morality perspective. *Behaviour & Information Technology*, *41*(3), 502–518. https://doi.org/10.1080/0144929X.2020.1818828

Van den Bos, K., Wilke, H. A., & Lind, E. A. (1998). When do we need procedural fairness? The role of trust in authority. *Journal of Personality and Social Psychology*, *75*(6), 1449–1458. https://doi.org/10.1037/0022-3514.75.6.1449

Verma, S., & Rubin, J. (2018). Fairness definitions explained. In 2018 IEEE/ACM International Workshop on Software Fairness (Fairware) (pp. 1–7). IEEE. https://doi.org/10.1145/3194770.3194776

Vimalkumar, M., Gupta, A., Sharma, D., & Dwivedi, Y. (2021). Understanding the effect that task complexity has on automation potential and opacity: Implications for algorithmic fairness. *AIS Transactions on Human-Computer Interaction*, *13*(1), 104–129. https://doi.org/10.17705/1thci.00144

Walker, K., Croak, M. (2021). An update on our progress in responsible AI innovation. *Google Blog*. https://blog.google/technology/ai/update-our-progress-responsible-ai-innovation/

Wang, A. J. (2018). Procedural justice and risk-assessment algorithms. SSRN. https://ssrn.com/abstract=3170136 or https://doi.org/10.2139/ssrn.3170136

Wang, R., Harper, F. M., Zhu, H. (2020, April). Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). ACM.

Wang, X., & Yin, M. (2021). Are explanations helpful? A comparative study of the effects of explanations in AI-assisted decision-making. In 26th International Conference on Intelligent User Interfaces (pp. 318–28). ACM. https://doi.org/10.1145/3397481.3450650

Wangmo, T., Lipps, M., Kressig, R. W., & Ienca, M. (2019). Ethical concerns with the use of intelligent assistive technology: Findings from a qualitative study with professional stakeholders. *BMC Medical Ethics*, *20*(1), 1–11. https://doi.org/10.1186/s12910-019-0437-z

Werth, R. (2019). Risk and punishment: The recent history and uncertain future of actuarial, algorithmic, and "evidence-based" penal techniques. *Sociology Compass*, *13*(2), e12659. https://doi.org/10.1111/soc4.12659

Wiener, M., Cram, W., & Benlian, A. (2021). Algorithmic control and gig workers: A legitimacy perspective of Uber drivers. *European Journal of Information Systems*. Advance online publication. https://doi.org/10.1080/0960085X.2021.1977729

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, *96*(4), 114–123. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

Wojcieszak, M., Thakur, A., Ferreira Gonçalves, J. F., Casas, A., Menchen-Trevino, E., & Boon, M. (2021). Can AI enhance people's support for online moderation and their openness to dissimilar political views? *Journal of Computer-Mediated Communication*, *26*(4), 223–243. https://doi.org/10.1093/jcmc/zmab006

Wonseok, J., Woo, K. Y., & Yeonheung, K. (2021). Who made the decisions: Human or robot umpires? The effects of anthropomorphism on perceptions toward robot umpires. *Telematics and Informatics*, *64*, 101695. https://doi.org/10.1016/j.tele.2021.101695

Woodruff, A., Fox, S. E., Rousso-Schindler, S., & Warshaw, J. (2018, April). A qualitative exploration of perceptions of algorithmic fairness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). ACM. https://doi.org/10.1145/3173574.3174230

Yalcin, G., Lim, S., Puntoni, S., & van Osselaer, S. M. (2022). Thumbs up or down: Consumer reactions to decisions by algorithms versus humans. *Journal of Marketing Research, 59*(4), 696–717. https://doi.org/10.1177/00222437211070016

Yang, S. J. H., Ogata, H., Matsui, T., & Chen, N.-S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*, *2*, 100008. https://doi.org/10.1016/j.caeai.2021.100008

Zahedi, Z., Sengupta, S., & Kambhampati, S. (2020). Why didn't you allocate this task to them?'Negotiation-Aware Task Allocation and Contrastive Explanation Generation. *arXiv preprint arXiv:2002.01640*

Zhang, L., & Yencha, C. (2022). Examining perceptions towards hiring algorithms. *Technology in Society*, *68*(C), 101848. https://doi.org/10.1016/j.techsoc.2021.101848

Zhang, P., Nah, F. F. H., & Preece, J. (2004). Guest editorial: HCI studies in management information systems. *Behaviour & Information Technology*, *23*(3), 147–151. https://doi.org/10.1080/01449290410001669905

Zhou, J., Verma, S., Mittal M.., & F., Chen. (2021). Understanding relations between perception of fairness and trust in algorithmic decision making. In 2021 8th International Conference on Behavioral and Social Computing (BESC) (pp. 1–5). ACM. https://doi.org/10.1109/BESC53957.2021.9635182

About the authors

Devesh Narayanan is a research assistant at the NUS Centre on AI Technology for Humankind. His research concerns the normative-theoretical and behavioural underpinnings of popular calls for "ethical", "trustworthy", and "human-centered" AI. He holds an M.A. in Philosophy and B.Eng. in Mechanical Engineering, both from the National University of Singapore.

Mahak Nagpal is a Postdoctoral Fellow at the Centre on AI Technology for Humankind, National University of Singapore (NUS) Business School. Broadly, her research considers ethical perspectives related to human-AI interaction in the workplace. She received her Ph.D. in Organization Management from Rutgers Business School.

Jack McGuire is a PhD candidate in the Department of Management & Organisation at NUS Business School. Prior to this, he was the Experimental Lab Manager of the Cambridge Experimental and Behavioural Economics Group (CEBEG) at Judge Business School, University of Cambridge.

Shane Schweitzer is a postdoctoral research associate in the Centre for Trusted Internet and Community and the Centre on AI Technology for Humankind at National University of Singapore. His current research concerns perceptions of advanced, humanlike technologies.

David De Cremer is a Provost's Chair and Professor of Management and Organization at the NUS Business School, and the Director of the Centre on AI Technology for Humankind. Before moving to NUS, he was the KPMG Endowed Professor of Management Studies at the Judge Business School, Cambridge University.