

Algorithmic Management and the Erosion of Managerial Trust: A Conceptual Framework for Human-AI Supervisory Relations in Organizations

 Aprajita Sen^{1*}
 Vishwas Kulkarni²
 Madanlal Tembhre³
 Mahesh Rawat⁴

How to cite this article:

Sen, A., Kulkarni, V., Tembhre, M., & Rawat, M. (2025). Algorithmic Management and the Erosion of Managerial Trust: A Conceptual Framework for Human-AI Supervisory Relations in Organizations. *Journal of Excellence in Management Sciences*, 3(1), 65–78.

Received: 1 January 2025 / Accepted: 9 May 2025 / Published online: 20 June 2025

Abstract

Supervising, monitoring, scheduling, evaluating performance, and sometimes even disciplining, are becoming increasingly delegated to algorithmic systems. The primary reason for implementing algorithmic management is to enhance efficiency, consistency, and scalability in managing employees. The paper claims that the same systems can, at the same time, take away the trust of managers, a different and not well-studied cost of an organization, which disproportionately affects the least powerful actors in the organization: workers. The paper is based on Algorithmic Management Theory, Psychological Contract Theory and Organizational Justice Theory and integrates the characteristics of algorithmic management, the opacity of algorithms, the intensity of surveillance, automated evaluation and constrained worker voice with a mediating process of managerial trust erosion, which consists of procedural injustice, interactional injustice and psychological contract breach. The framework identifies the factor of organizational outcomes as turnover intention, discretionary effort, employee voice, and organizational legitimacy as responsible for this erosion. Moderators include algorithmic transparency, human-override discretion, worker voice mechanisms, algorithmic literacy, justice climate and occupational context. Eight testable propositions are presented and the framework is expanded to hybrid, white-collar knowledge work, where algorithmic supervision is becoming ubiquitous with the guise of workplace analytics and productivity software. The paper ends by outlining a research agenda for developing algorithmic supervision without undermining the trust that sustains the authority of management.

Keywords: Algorithmic management; managerial trust; psychological contract; organizational justice; algorithmic control; employee voice; human resource management;

¹ Department of Management, IES College of Technology, Kalkheda, Ratibad Main Road, Bhopal-462044, Madhya Pradesh, India.

Corresponding Author: senaprajita631@gmail.com

² Beijing No.80 High School, 2 Baijiazhuang Rd, Chaoyang, China

³ Department of Psychology, D.P. Chaturvedi College of Science, Commerce, Arts and Education, Seoni, Madhya Pradesh, India.

⁴ Department of Commerce & Finance, IES University, Ratibad, Main Rd, Bhopal-462044, Madhya Pradesh, India.



future of work

1 Introduction

In most workplaces, organizations in all industries have outsourced certain supervisory responsibilities, task delegation, performance monitoring, scheduling, assessment and enforcement, to algorithmic systems (Kellogg et al., [2020](#)). The feature of platform-mediated gig work, ride-hailing dispatch, food delivery routing, freelance-marketplace rating, now has spread to the traditional world of work in the form of workplace analytics software, automated performance-management dashboards, and AI-assisted hiring and scheduling tools (Parent-Rocheleau & Parker, [2022](#)). The reason why employers use these systems is that they offer the promise of consistency, scalability, and less discretionary bias that affects human supervisors (Möhlmann et al., [2021](#)). Similar trends toward data-driven and algorithmic supervision in platform labour and human resource management (HRM) were reported in related research (Duggan et al., [2020](#); Lee et al., [2015](#); Tambe et al., [2019](#)).

This efficiency story is complicated by research into algorithmic management. Kellogg, Valentine, and Christin ([2020](#)) compile the multidisciplinary literature on the use of algorithms in the workplace and classify the various ways in which they do so into six mechanisms: controlling, restricting, recommending, recording, rating, replacing, and rewarding workers, which are increasingly mediated by the algorithms themselves. These mechanisms are not merely routine and repetitive components of managerial practices; they transform these practices, substituting for situated managerial judgment the manifestation of rules and scores that are sometimes incomprehensible for the worker being managed (Rahman, [2021](#)). When workers are evaluated using opaque algorithms, it is difficult to connect the work to the criteria by which it will be judged, and will create a unique form of disempowerment that Rahman ([2021](#)) calls an invisible cage. Investigations of online platforms and ride-hailing reveal other avenues by which information asymmetries and algorithmic control can limit the freedom and agency of workers (Curchod et al., [2020](#); Rosenblat & Stark, [2016](#); Wood et al., [2019](#)).

This paper suggests that one of the resources that can be affected by algorithmic management but is rarely measured first-hand is managerial trust, which is defined as the employees' trust that management — however represented — will treat them fairly and honor their implicit commitments under the employment contract (Rousseau, [1995](#)). Psychological contract theory suggests that when employers behave ambiguously, the employees' inferences about employer promises lead to disproportionate reactions of outrage and withdrawal, depending on the subjective severity of the violation based on the employee's expectations (Robinson, [1996](#)). Algorithmic management is a potential source of such violations since it can often strip away the explanatory, discretionary, and relational aspects that have traditionally supported workers' perception of a manager's decision as legitimate (Newman et al., [2020](#)). This is consistent with the findings that trust in leaders and organisational authorities are linked to important employee attitudes and behaviours at multiple levels of analysis (Dirks & Ferrin, [2002](#); Fulmer & Gelfand, [2012](#)).

This shift has stakes, as demonstrated by empirical studies showing that transparency of algorithmic management relates directly to organizational attitudes. Perceived transparency of the algorithmic surveillance and performance management was correlated with greater perceived distributive justice and procedural justice and less intention to quit among truck drivers who were algorithmically monitored and managed (Bujold et al., [2022](#)). Most deployed algorithmic management systems, however, are being bought for efficiency and compliance reasons, in which the role of explaining to workers is marginalized or neglected altogether (Parent-Rocheleau & Parker, [2022](#)). The empirical foundation for this argument here is the speed disparity between the speed at which it is possible to diffuse the algorithmic supervision and the comparatively slower pace of trust-preserving design practice. The studies in the field of eMonitoring also demonstrate

that perceived control, transparency, privacy, and monitoring design impact employee acceptance and reactions to justice (Abraham et al., 2019; Alge, 2001; Ravid et al., 2020).

There are three contributions to this paper. First, it draws the concept of Algorithmic Management Theory (Kellogg et al., 2020), Psychological Contract Theory (Rousseau, 1995; Robinson, 1996) and Organizational Justice Theory (Colquitt, 2001) within a single framework to explain the production and propagation of managerial trust erosion due to algorithmic supervision. Second, it highlights the factors that can serve as moderators between algorithmic management and the outcomes of eroded trust or a measured offset by institutional design (Bujold et al., 2022; Möhlmann et al., 2021). Third, it develops eight testable propositions and extends the platform-centric paradigm to the hybrid, white-collar knowledge workplace, where algorithmic supervisory tools are becoming a part of everyday corporate software (Parent-Rocheleau & Parker, 2022).

The rest of the paper summarizes the three theoretical foundations, discusses the conceptual model, presents the gap in the literature that the conceptual model fills, extends the argument to knowledge work, talks about implications for theory and practice, and concludes.

Figure 1. Conceptual Framework of Algorithmic Managerial Trust Erosion

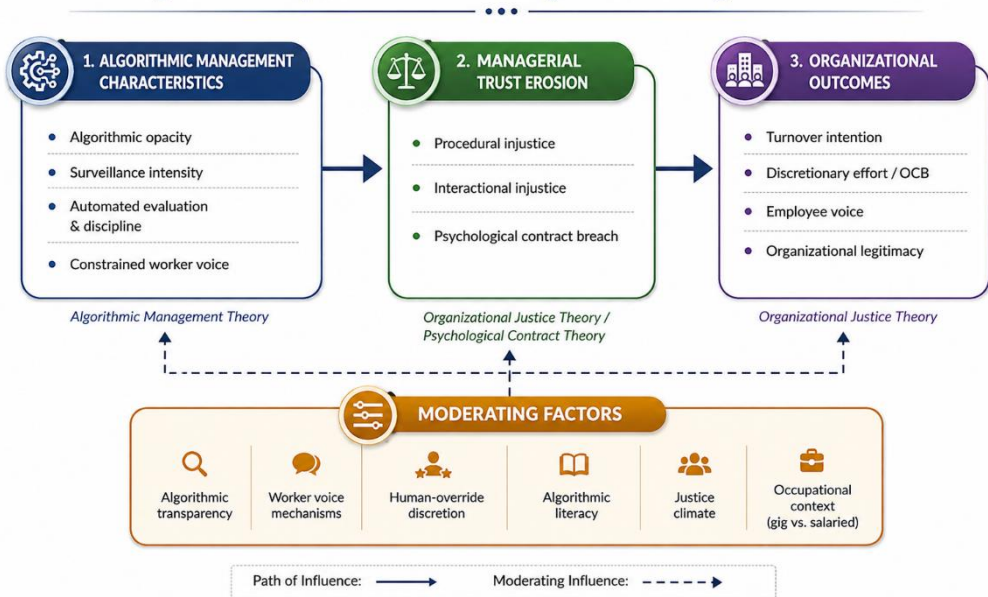


Figure 1. The framework traces a path from algorithmic management characteristics, through managerial trust erosion, to organizational outcomes, with six factors moderating the strength of these relationships.

2 Theoretical Foundations

The framework draws on the following three theories in organizational behavior and management: Algorithmic Management Theory (Kellogg et al., 2020); Psychological Contract Theory (Rousseau, 1995, Robinson, 1996); and Organizational Justice Theory (Colquitt, 2001). These traditions highlight various aspects of the employee-supervisor relationship, and a combination of these traditions is required to theorize how this relationship is being transformed with algorithmic supervision (Newman et al., 2020).

2.1 Algorithmic Management Theory

Algorithmic Management Theory, developed by Kellogg, Valentine and Christin (2020), is based on Edwards' framework of labor process, a contested terrain in which the employer uses a production technology to squeeze labour value, and the worker adapts or resists by resisting. Kellogg et al. (2020) note six mechanisms, the "6 Rs", by which algorithms wield workplace control: directing workers through restriction and recommendation; evaluating workers through recording and rating; replacing workers and disciplining workers with replacement; rewards workers through rewarding. These processes are examples of rational control as opposed to the technical and bureaucratic control of the twentieth century (Kellogg et al., 2020).

Algorithmic control is not limited to gig platforms. Möhlmann, Zalmanson, Henfridsson and Gregory (2021) differentiate between algorithmic matching (the process of connecting labour supply and demand) and algorithmic control (the constant supervision, coaxing and assessment of workers who have been matched with a job) and note that algorithmic control is increasingly becoming like "traditional managerial power exercised through code" (Möhlmann et al., 2021). Parent-Rochelleau and Parker (2022) extend this study to six different management functions: monitoring, goal setting, performance management, scheduling, compensation, and termination—functions that, when handled by algorithms, are more likely to increase the intensity of work and decrease workers' autonomy, since they are not able to communicate with the algorithm or ask questions about it as they would a human manager.

One of the central empirical findings in this literature is that algorithmic opacity is a cause of worker harm— apart from the substantive decisions an algorithm makes. Although Rahman's (2021) multi-year study of a freelance labour platform did not examine Algieri, her work revealed a lack of alignment between a worker's effort and its evaluation algorithm because the algorithm was opaque, meaning that workers were unable to know its criteria, and were not provided with any behavioural guidance, resulting in chronic anxiety and a metaphorical invisible cage of behavioural constraint. In this paper, algorithmic opacity, surveillance intensity, automated assessment and discipline, and limited worker voice are defined as the specific forms of the antecedents of managerial trust erosion that are brought to the fore by AI. In general, the studies on workplace surveillance have been associated with privacy concerns and negative attitudes toward work due to intensive monitoring, particularly in cases where such monitoring is perceived as intrusive or tightly controlling (Ball, 2010; Jeske & Santuzzi, 2015).

2.2 Psychological Contract Theory

According to the Psychological Contract Theory developed by Rousseau (1995), workers develop their own understandings of the mutual commitments of the employment relationship through the promises or expectations they receive, either explicitly or in the form of implicit signals from their organization, recruitment communications, and social norms. The psychological contract is a subjective perception and not a written agreement, hence it exists in the eye of the beholder, and is not a disappointment, but rather a distinct emotional event when it is violated.

Robinson's (1996) longitudinal study of newly hired managers revealed that the relationship between psychological contract breach and the subsequent development of employee trust is strong and bidirectional: having higher levels of trust at the start of the relationship increased the probability that employees will experience psychological contract breach, and vice versa. Robinson (1996) also demonstrates that, when a manager's actions are ambiguous, employees with high trust are more likely to see a violation where managers are using an algorithmic decision, than employees with low trust are likely to see a violation for the same manager using the identical algorithmic decision. The effects of psychological contract breach have been associated with lower levels of organizational citizenship behavior (Robinson, 1996), decreased job satisfaction, and, in extreme cases, retaliation against the employer (Robinson, 1996).

In the context of algorithmic management, a psychological contract suggests that employees who

were hoping one of their supervisors would explain a decision, have discretion in a difficult case, or simply listen, will feel that they have not fulfilled their contractual obligations when the decision is made entirely through an algorithm, and no supervisor was explicitly promised (Newman et al., [2020](#)). In this paper, psychological contract breach is one of three aspects of the managerial trust erosion that algorithmic management can create.

2.3 Organizational Justice Theory

According to Organizational Justice Theory (Colquitt, [2001](#)), Organizational Justice is divided into four distinct dimensions: Distributive Justice, how they perceive the outcome; Procedural Justice, how they perceive the process that leads to the outcome; Interpersonal Justice, how they perceive the communication/feelings of the process; and Informational Justice, how they perceive the information from the process. Confirmatory factor analyses conducted at the University and in a manufacturing firm sample by Colquitt ([2001](#)) validated that these dimensions are empirically separate and that they have differential impacts on outcome variables such as evaluation of the leader, rule compliance, organizational commitment, and helping behavior. A further body of evidence, from a meta-analysis, supports the importance of justice perceptions for social-exchange and affective outcomes in organizations (Colquitt et al., [2013](#)).

Even if an algorithm removes the biases that would be present in human decisions, when employees know that HR decisions are made by an algorithm, they perceive them as less procedurally just than when they are made by humans, since algorithmic decision making is perceived to decrease individualized consideration and voice, both antecedents of procedural justice. This is significant as it demonstrates that perceived fairness in a decision is not only dependent on removing bias, but is also dependent on the process used to achieve the decision (Newman et al., [2020](#)). Another central concern of implementation identified in a systematic review of algorithmic HR decision-making is perceived unfairness and discrimination risks (Köchling & Wehner, [2020](#)).

It has been confirmed in the field. In algorithmic managed truck drivers, the perceived transparency of the algorithmic management was significantly positively correlated with distributive justice and procedural justice, which were also significantly positively correlated with reduced intention to leave the organization (Bujold et al., [2022](#)). This paper focuses on procedural and interactional injustice as the second and third dimensions of the breach of managerial trust, in addition to the psychological contract, and uses Colquitt's ([2001](#)) justice dimensions as the main lens for the outcome side of the framework.

3 Methodological Approach to Framework Development

This paper adopts the convention for conceptual and theory building papers in organizational behavior and management: instead of testing hypotheses based on new data, the paper brings together theoretical traditions and proposes a new mediating construct, managerial trust erosion, and a set of testable relations that link this construct to its moderators, antecedents, and outcomes. The framework was developed by an integrative review of three streams of literature largely developed independently from each other: algorithmic management literature (spanning Kellogg et al. [2020](#), with subsequent extensions by Möhlmann et al. [2021](#), Parent-Rochelleau and Parker [2022](#), and Rahman [2021](#)); psychological contract theory (spanning Rousseau [1995](#), and extending by Robinson [1996](#)); and organizational justice theory (spanning Colquitt [2001](#) and applied to algorithmic contexts by Newman et al. [2020](#) and Bujold et al. [2022](#)).

Constructs were selected on the basis of having strong theoretical foundations in at least one of the foundational traditions (Kellogg et al., [2020](#); Rousseau, [1995](#); Colquitt, [2001](#)) and empirical referents in the applied algorithmic-management literature (Rahman, [2021](#); Bujold et al., [2022](#); Möhlmann et al., [2021](#)) to ensure the framework is recognized by, and testable with, researchers

in real deployments of algorithmic supervision rather than just hypothetical ones. The resulting framework identifies four antecedents, three dimensions of trust erosion, four outcomes, and six moderators, the number that was selected to be a manageable number to balance comprehensiveness with feasibility for future empirical research.

3.1 A Conceptual Framework of Algorithmic Managerial Trust Erosion

Building upon the three traditions reviewed above, this paper proposes a framework with four interconnected constructs: algorithmic management characteristics as antecedents (Kellogg et al., [2020](#)); managerial trust erosion, which includes three subconstructs, namely procedural injustice, interactional injustice, and psychological contract breach (Colquitt, [2001](#); Rousseau, [1995](#)); organizational outcomes, which are turnover intention, discretionary effort, employee voice, and organizational legitimacy (Bujold et al., [2022](#); Parent-Rocheleau & Parker, [2022](#)); and moderating factors that condition the strength of these relationships (Möhlmann et al., [2021](#)). The framework is summarized in Figure 1 and in the propositions which follow.

3.2 Antecedents: Algorithmic Management Characteristics

Four features of algorithmic management are suggested as factors that contribute to erosion of managerial trust.

The opacity of algorithms indicates the extent to which the principles and processes used in an algorithm rating are not accessible or comprehensible to the person being rated (Rahman, [2021](#)). There are many possible causes of opacity, including technical complexity, deliberate opacity to prevent workers from "gaming" an evaluation, and failure to invest in explanation to workers, which all create a unique form of disempowerment in which the worker is unable to match his or her actions with what is used to judge him or her (Rahman, [2021](#)).

Surveillance intensity is the degree and granularity of algorithmic tracking of the worker's performance, where, how fast, and what the worker is producing (Kellogg et al., [2020](#)). Productivity tracking and remote monitoring software are other forms of algorithmic surveillance that are found in traditional workplaces, and the intensity of this form of algorithmic management is among the two algorithmic management functions that were found to affect perceptions of justice among workers who were subjected to algorithmic management. (Bujold et al., [2022](#))

Automated evaluation and discipline can be characterized by the extent to which performance rating, warning, or termination decisions are made by an algorithm, in the absence of meaningful human involvement (Kellogg et al., [2020](#)). Despite such measures to minimize the discriminatory bias of human evaluators, workers see algorithmic-based evaluations as less procedurally fair because algorithmic evaluations are linked to lower individualized consideration (Newman et al., [2020](#)).

Constrained worker voice is defined as how much space a worker has to question, challenge, or get clarification on an algorithmic decision (Parent-Rocheleau & Parker, [2022](#)). Algorithmic management actions like compensation and scheduling often occur in isolation of a communicative back channel to the worker, increasing demands and blocking any possibility of negotiation that a human scheduler might have allowed (Parent-Rocheleau & Parker, [2022](#)).

3.3 Managerial Trust Erosion

They are hypothesized to contribute to the erosion of managerial trust on three dimensions based on the theories of Organizational Justice Theory and Psychological Contract Theory.

Procedural injustice is when a worker perceives that the way that an algorithmic decision was made was not neutral, transparent, or consistent (Colquitt, [2001](#)). Algorithmic decision-making is seen as less procedurally fair than human decision making, even when the decision is not only

accurate, but also free of bias (Newman et al., [2020](#)).

Workers' sense that they were not treated with dignity, respect, or proper explanation during an algorithmic decision is referred to as interactional injustice following the work of Colquitt ([2001](#)). Without a human interlocutor, an algorithm can make a decision but cannot provide an explanation as interpersonal justice provides, thus making structurally likely a certain degree of interactional injustice to occur (Rahman, [2021](#)).

Psychological contract breach is defined as when employees feel that the organisation has violated an implicit employment contract, such as when they feel that they were not treated with fairness, human consideration, or were not given an explanation for the actions taken (Rousseau, [1995](#)). Since trust is a lens, workers who already distrust their employer are more likely to interpret an ambiguous algorithmic decision as a violation, whereas workers whose trust is high will give the organization the benefit of the doubt – at least until there are repeated ambiguous algorithmic decisions (Robinson, [1996](#)).

3.4 Organizational Outcomes

Once created, managerial trust erosion is then suggested to spread to four organizational outcomes that are central to both the justice literature and the psychological-contract literature.

The turnover intention outcome is the one most directly evidenced: Among algorithmically managed truck drivers, increased perceived transparency of the surveillance system and increased perceived justice mediated this pathway, showing that the link between perceived transparency and retention is mediated by perceived justice (Bujold et al., [2022](#)).

Trust erosion is hypothesized to be negatively associated with discretionary effort and organizational citizenship behavior as evidence suggests that when trust is breached in the psychological contract, employees will be less motivated to give the organization more than they are required by their jobs (Robinson, [1996](#)).

Theoretically, employee voice decreases because when an algorithm is used to manage an organization, there is often no communicative avenue back to the employee, and when employees learn that they can't petition the algorithm, they may also not be willing to petition the organization (Parent-Rochelleau & Parker, [2022](#)). A wider range of voice research reveals that the support of employees in speaking up is greatly influenced by whether they believe that the voice is genuinely accepted by the organization (Detert & Burris, [2007](#); Morrison, [2023](#)).

Organizational legitimacy is theorized to suffer as each of the three outcomes above stack up across the workforce as control through algorithmic processes is opaque, unexplained, and unilateral, making them inconsistent with the interactional norms that underlie perceptions of legitimate authority (Kellogg et al., [2020](#)). Suchman ([1995](#)) and Tyler ([2006](#)) emphasized perceptions of appropriateness, fairness, and justified authority in organizational and psychological accounts of legitimacy, which is congruent with this reasoning.

3.5 Moderating Factors

It outlines six factors that moderate the relationship between algorithmic management characteristics to managerial trust erosion, and the relationship between trust erosion to adverse organizational outcomes.

Algorithmic transparency is the extent to which the workers are provided with an intelligible explanation or description of how the system monitors, evaluates, or scores them. Greater perception of distributive justice and procedural justice among algorithmically managed workers is directly linked to algorithmic transparency of both algorithmic surveillance and algorithmic performance management (Bujold et al., [2022](#)).

The human-override discretion questions whether there is useful discretion available to a human supervisor to review, explain, and/or overrule an algorithmic recommendation prior to its impact on a worker. Algorithmic control that fully replaces this discretion takes away the interactional and explanatory tools that workers have traditionally used to make a decision feel legitimate (Kellogg et al., [2020](#)). This is also a sign of the automation-augmentation paradigm in AI-assisted management, in which combining meaningful human judgment with automated operations can limit the drawbacks of complete automation (Raisch & Krakowski, [2021](#)).

Worker voice mechanisms are the avenues through which workers can challenge or explain an algorithmic decision. Scheduling and compensation via algorithms further increase the demands of work in comparison to their human counterparts, one of the main reasons being the lack of scheduling channels (Parent-Rochelleau & Parker, [2022](#)).

Algorithmic literacy is the individual's ability to understand the algorithms underlying its operation. An opaqueness that prevents workers from understanding a score or notice they don't know, is a disempowering factor, as they won't be able to correlate their behavior with evaluation criteria (Rahman, [2021](#)).

Justice climate is the general organisational context in which an algorithmic system is situated, comprising of distributive justice, procedural justice and interactional justice (Colquitt, [2001](#)). A justice climate of an organization with otherwise high levels of trust might mitigate the impact of a particular algorithm that is opaque, but a low justice climate could amplify the negative impact.

Occupational context is the type of work that algorithmic management is used in: Gig economy/platform-mediated work or traditional, salaried jobs. Whether workers have other ways to earn income and exit from the algorithmic control differs, and in the latter case, the algorithmic control is experienced as domination instead of merely as a structure, as Möhlmann et al. ([2021](#)) describe as central to understanding of algorithmic control as domination.

Table 1 summarizes the eight propositions that follow from this framework, linking each construct pairing to its theoretical foundation.

Table 1

Propositions Linking the Constructs of the Framework

#	Proposition	Primary Theoretical Basis
P1	Algorithmic opacity is positively associated with workers' perceived procedural injustice.	Organizational Justice Theory
P2	Automated evaluation and discipline without human review is positively associated with perceived interactional injustice.	Organizational Justice Theory
P3	Constrained worker voice in algorithmic decision-making is positively associated with psychological contract breach.	Psychological Contract Theory
P4	Managerial trust erosion is positively associated with turnover intention.	Psychological Contract Theory; Organizational Justice Theory
P5	Managerial trust erosion is negatively associated with discretionary effort and organizational citizenship behavior.	Psychological Contract Theory
P6	Algorithmic transparency and human-override discretion weaken the negative relationship between algorithmic management and	Algorithmic Management Theory

	organizational outcomes.	
P7	Worker voice mechanisms are positively associated with perceived procedural justice under algorithmic management.	Organizational Justice Theory
P8	Perceived justice under algorithmic management is positively associated with organizational legitimacy.	Organizational Justice Theory; Algorithmic Management Theory

Table 2
Summary of Framework Constructs

Framework Component	Elements
Antecedents	Algorithmic opacity; surveillance intensity; automated evaluation and discipline; constrained worker voice
Managerial trust erosion	Procedural injustice; interactional injustice; psychological contract breach
Organizational outcomes	Turnover intention; discretionary effort and organizational citizenship behavior; employee voice; organizational legitimacy
Moderating factors	Algorithmic transparency; human-override discretion; worker voice mechanisms; algorithmic literacy; justice climate; occupational context

3.6 Research Gap

Recent scholarship on algorithmic management has advanced understanding of its mechanisms of Research on algorithmic management has increased the understanding of how algorithms manage or control work (Kellogg et al., 2020), what effects algorithms have on job design and autonomy (Parent-Rochelleau & Parker, 2022), and the lived experience of algorithmic evaluation (Rahman, 2021). A systematic account of how algorithmic supervision undermines managerial trust as a valuable organizational resource, and not an undifferentiated side effect of automation, is underdeveloped. In most studies that investigate attitudes towards algos (Bujold et al., 2022), attitudes are measured directly, and the psychological mechanism, breach of an implicit relational obligation, that links a characteristic of the algos to that attitude is not explicitly theorized. Algorithmic management is not only a matter of technology but also of the broader implications for employment relations that can spill over from individual decisions. This is backed up by related reviews and field studies (Duggan et al., 2020; Wood et al., 2019).

This gap is significant because of three reasons. First, algorithmic management research has largely grown out of labor process theory and platform-work studies (Kellogg et al., 2020; Möhlmann et al., 2021); the extension of algorithmic management constructs into mainstream constructs in organizational behavior, e.g., the psychological contract, has not been systematically developed despite their potential to explain the disproportionate reactions to algorithmic decisions, which are only as severe as the actual situation demands. (Robinson, 1996). Second, organizational justice studies of algorithmic decisions have been about isolated decision episodes (Newman et al., 2020), a single hiring or evaluation decision, and not about repeated or cumulative exposure as is the case in ongoing algorithmic supervision. Third, evidence that transparency effects increase justice perceptions and retention exists for specific occupational contexts (e.g., trucking; Bujold et al., 2022), but a framework exists that specifies when such effects should generalize across occupational contexts is not yet available.

This paper's theoretical foundation is designed to fill this gap by providing a theorized testable explanation for when algorithmic management is likely to uphold managerial trust and when it is likely to create a relational breach that will result in disproportionately high organizational costs (Rousseau, [1995](#); Colquitt, [2001](#)).

Enabling the Framework: Platform Gig Work to Hybrid Knowledge Workplace

The field of algorithmic management research began primarily through investigations of platform-mediated gig work, ride-hailing, food delivery, and freelance marketplaces, where the use of algorithms was most pronounced and most pervasive (Möhlmann et al., [2021](#); Rahman, [2021](#)). This framework, however, is becoming more applicable to traditional paid work, as algorithms are increasingly incorporated into standard workplace software, for example, under the guise of workplace analytics, productivity monitoring and AI-enabled performance management (Parent-Rocheleau & Parker, [2022](#)).

The framework lists antecedents that are becoming more common in hybrid and remote white-collar work that can contribute to the loss of managerial trust. When it comes to surveillance, those numbers have jumped considerably since the introduction of remote-work monitoring software that logs what applications workers are using, how long they are working, and what they are typing, functions that are strikingly similar to the algorithmic surveillance mechanism found among gig and platform workers (Kellogg et al., [2020](#)). AI is also making its way into knowledge work via performance review software that uses AI to summarise or rate employee work, sometimes without the employee rated even knowing what the criteria were to achieve a particular score (Rahman, [2021](#)). Such issues are also becoming more relevant when it comes to systems that use Artificial Intelligence (AI) for HR (Ravid et al., [2020](#); Tambe et al., [2019](#)), as these are becoming widely diffused beyond platform labor.

The difference between salaried knowledge workers and gig workers is important, though, and one that the moderators of the framework should reflect. Occupational context is important because there is at least a formal human manager named in the contract, which may allow for the continuation of the human 'override' that gig platforms do not always include (Möhlmann et al., [2021](#)). This formal authority could be replacing direct human review, or it could be a symbolic gesture in order to make a decision that the algorithm had already made. Either way, the empirical question of whether managerial trust erosion in knowledge work resonates with or varies from the gig-economy pattern (Kellogg et al., [2020](#)) is open, and with implications for whether or not this form of authority is truly effective.

Concurrently, psychological contracts of salaried workers may be more entrenched and definite than those of gig workers, who tend to have more transactional relationships with a platform (Rousseau, [1995](#)). If this is the case, the same algorithmic attribute, such as a vague productivity measure, might trigger a much more intense reaction of the conflict in a long-term employee receiving a salary compared to a gig worker who has lower expectations of organizational reciprocity (Robinson, [1996](#)). This indicates that algorithmic managerial trust erosion may not be linearly transferred from gig to knowledge-work contexts and that occupational context needs to be explicitly studied as an empirical variable rather than a background control variable (Parent-Rocheleau & Parker, [2022](#)).

4 Discussion and Implications

This paper's central theoretical contribution is to show that Algorithmic Management Theory, Psychological Contract Theory, and Organizational Justice Theory are mutually necessary for understanding algorithmic supervision. Algorithmic Management Theory supplies the vocabulary for the control mechanisms at issue (Kellogg et al., [2020](#)); Organizational Justice Theory supplies the criteria by which workers judge those mechanisms fair or unfair (Colquitt, [2001](#)); and

Psychological Contract Theory explains why a judgment of unfairness can escalate into a disproportionate relational breach (Rousseau, [1995](#); Robinson, [1996](#)). No single tradition alone captures the phenomenon: Algorithmic Management Theory alone would describe the mechanisms of control without theorizing their downstream relational costs; Organizational Justice Theory alone would supply evaluative criteria without an account of why algorithmic decisions are judged less fair even when less biased (Newman et al., [2020](#)); and Psychological Contract Theory alone would explain the intensity of workers' reactions without specifying which algorithmic characteristics trigger them.

For practice, the moderating factors identified here, transparency, human-override discretion, worker voice, algorithmic literacy, justice climate, and occupational context, are conditions for the legitimate use of algorithmic supervision rather than optional enhancements (Bujold et al., [2022](#)). Organizations that adopt algorithmic management primarily to reduce supervisory headcount or standardize evaluation, without parallel investment in these moderators, are likely to realize efficiency gains that are partly illusory once elevated turnover and reduced discretionary effort are accounted for (Robinson, [1996](#); Parent-Rocheleau & Parker, [2022](#)). This reframes algorithmic-management adoption as a trust-allocation decision with direct consequences for retention and legitimacy, not merely a workforce-technology procurement decision (Kellogg et al., [2020](#)). In practice, monitoring systems are more likely to be accepted when employees perceive greater control and when monitoring is implemented with attention to privacy and fairness (Abraham et al., [2019](#); Alge, [2001](#)).

For researchers, the framework offers a structured basis for empirical work that has so far been fragmented across labor process studies, HRM technology adoption research, and organizational justice research (Kellogg et al., [2020](#); Colquitt, [2001](#)). The eight propositions are testable using field surveys of the kind already used to study algorithmic transparency and justice among truck drivers (Bujold et al., [2022](#)), vignette experiments of the kind used to isolate algorithmic reductionism effects in HR decisions (Newman et al., [2020](#)), or longitudinal designs of the kind Robinson ([1996](#)) used to trace the co-evolution of trust and perceived breach over time. The framework can also be tested against established evidence on leadership trust, organizational justice, and employee voice to assess whether algorithmic supervision changes the strength or form of these relationships (Colquitt et al., [2013](#); Dirks & Ferrin, [2002](#); Morrison, [2023](#)).

4.1 Limitations and Future Research

This is a conceptual paper and the propositions are intended as hypotheses for future testing research (Kellogg et al., [2020](#)). The framework also streamlines a diverse set of algorithmic management applications, including scheduling algorithms, performance-scoring systems, and termination recommendations powered by AI, which vary significantly in terms of stakes and reversibility; future studies should investigate whether the constructs of the framework differ across these subtypes (Parent-Rocheleau & Parker, [2022](#)). The moderating variables are assumed to be largely independent, but there is likely some interaction, such as the role of worker voice mechanisms depending on algorithmic literacy, because a contest channel would be ineffective for a worker who is unable to understand the score being contested (Rahman, [2021](#)). If the propositions made here are to be more robustly supported by external validity, then systematic comparisons between the different occupational contexts, gig, hybrid, and traditional salaried work, would need to be made (Möhlmann et al., [2021](#)). Future studies could focus on mixed methods studies which incorporate organizational turnover and engagement data, employee surveys that measure the three dimensions of trust erosion and four outcomes, and qualitative interviews that capture workers' lived experience of algorithmic supervision (Rahman, [2021](#); Bujold et al., [2022](#)).

4.2 Conclusion

The design practices to ensure trust are being implemented more slowly than algorithmic

supervision in the organizations (Kellogg et al., 2020). This paper has been about the possibility of the efficiency that's supposed to be the hallmark of algorithmic management existing alongside, and indeed creating, decreases in managerial trust, which are disproportionately felt by workers who are least able to challenge the opaque decision (Rahman, 2021; Newman et al., 2020). The theory of Algorithmic Management, the theory of Psychological Contract, and the theory of Organizational Justice Theory (Kellogg et al., 2020; Rousseau, 1995; Colquitt, 2001) are used together in this paper to create a framework that maps the pathway between the characteristics of algorithmic management, the erosion of managerial trust, and organizational outcomes, and that identifies which institutional conditions can help buffer the worst impacts of that pathway (Bujold et al., 2022). The propositions of this framework present a research agenda for scholars interested in organizational behaviour and human resource management, and the move from gig work to hybrid knowledge work suggests that the implications of algorithmic supervision are not limited to the platform economy in which it was first reported (Parent-Rochelleau & Parker, 2022). A central problem to be addressed is the balance between the efficiency of algorithmic supervision for the organization and the trust that is required to support its managerial authority.

Conflict of interest

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

Data availability

Data will be furnished upon request

5 References

- Abraham, M., Niessen, C., Schnabel, C., Lorek, K., Grimm, V., Möslin, K., & Wrede, M. (2019). Electronic monitoring at work: The role of attitudes, functions, and perceived control for the acceptance of tracking technologies. *Human Resource Management Journal*, 29(4), 657–675. <https://doi.org/10.1111/1748-8583.12250>
- Alge, B. J. (2001). Effects of computer surveillance on perceptions of privacy and procedural justice. *Journal of Applied Psychology*, 86(4), 797–804. <https://doi.org/10.1037/0021-9010.86.4.797>
- Ball, K. (2010). Workplace surveillance: An overview. *Labor History*, 51(1), 87–106. <https://doi.org/10.1080/00236561003654776>
- Bujold, A., Parent-Rochelleau, X., & Gaudet, M.-C. (2022). Opacity behind the wheel: The relationship between transparency of algorithmic management, justice perception, and intention to quit among truck drivers. *Computers in Human Behavior Reports*, 8, Article 100245. <https://doi.org/10.1016/j.chbr.2022.100245>
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, 86(3), 386–400. <https://doi.org/10.1037/0021-9010.86.3.386>
- 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. *Journal of Applied Psychology*, 98(2), 199–236. <https://doi.org/10.1037/a0031757>
- Curchod, C., Patriotta, G., Cohen, L., & Neysen, N. (2020). Working for an algorithm: Power asymmetries and agency in online work settings. *Administrative Science Quarterly*, 65(3), 644–676. <https://doi.org/10.1177/0001839219867024>
- open? Academy of Management Journal, 50(4), 869–884. <https://doi.org/10.5465/amj.2007.26279183>
- Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications

- for research and practice. *Journal of Applied Psychology*, 87(4), 611–628. <https://doi.org/10.1037/0021-9010.87.4.611>
- Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. *Human Resource Management Journal*, 30(1), 114–132. <https://doi.org/10.1111/1748-8583.12258>
- Fulmer, C. A., & Gelfand, M. J. (2012). At what level (and in whom) we trust: Trust across multiple organizational levels. *Journal of Management*, 38(4), 1167–1230. <https://doi.org/10.1177/0149206312439327>
- Jeske, D., & Santuzzi, A. (2015). Monitoring what and how: Psychological implications of electronic performance monitoring. *New Technology, Work and Employment*, 30(1), 62–78. <https://doi.org/10.1111/ntwe.12039>
- 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>
- Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795–848. <https://doi.org/10.1007/s40685-020-00134-w>
- Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1603–1612). Association for Computing Machinery. <https://doi.org/10.1145/2702123.2702548>
- Morrison, E. W. (2023). Employee voice and silence: Taking stock a decade later. *Annual Review of Organizational Psychology and Organizational Behavior*, 10, 79–107. <https://doi.org/10.1146/annurev-orgpsych-120920-054654>
- Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. (2021). Algorithmic management of work on online labor platforms: When matching meets control. *MIS Quarterly*, 45(4), 1999–2022. <https://doi.org/10.25300/MISQ/2021/15333>
- 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>
- Parent-Rocheleau, X., & Parker, S. K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. *Human Resource Management Review*, 32(3), Article 100838. <https://doi.org/10.1016/j.hrmr.2021.100838>
- Rahman, H. A. (2021). The invisible cage: Workers' reactivity to opaque algorithmic evaluations. *Administrative Science Quarterly*, 66(4), 945–988. <https://doi.org/10.1177/00018392211010118>
- 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>
- Ravid, D. M., Tomczak, D. L., White, J. C., & Behrend, T. S. (2020). EPM 20/20: A review, framework, and research agenda for electronic performance monitoring. *Journal of Management*, 46(1), 100–126. <https://doi.org/10.1177/0149206319869435>
- Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly*, 41(4), 574–599. <https://doi.org/10.2307/2393868>
- of Uber's drivers. *International Journal of Communication*, 10, 3758–3784. <https://doi.org/10.2139/ssrn.2686227>
- Rousseau, D. M. (1995). *Psychological contracts in organizations: Understanding written and unwritten agreements*. Sage Publications. <https://doi.org/10.4135/9781452231594>

- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571–610. <https://doi.org/10.5465/amr.1995.9508080331>
- Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15–42. <https://doi.org/10.1177/0008125619867910>
- Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology*, 57, 375–400. <https://doi.org/10.1146/annurev.psych.57.102904.190038>
- Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work, Employment and Society*, 33(1), 56–75. <https://doi.org/10.1177/0950017018785616>