Heimer Media

Review of Economic & Managerial Studies

www.heimer.cn/index.php/rems



The Theory Adaptation Approach to Trust in Algorithmic Management

Freya Hughes [1]*

1. University of Cumbria, UK
* Correspondence: freyahughes@cumbria.ac.uk

Abstract: Algorithmic management, e.g., the use of software and machine-learning systems to delegate, evaluate, and discipline work—has expanded rapidly across platform-mediated gig economies and, increasingly, conventional organizations. Despite a growing body of empirical work on worker perceptions of algorithmic control, the field lacks a coherent theoretical account of why workers do or do not place trust in the algorithmic systems that govern their livelihoods. This paper adopts the Theory Adaptation Approach, an established methodology in information systems research, to systematically adapt integrative model of organizational trust for the algorithmic management context. Following the reflexive guidelines for theory adaptation and the integrated adaptation guidelines, we examine the fit between the source trust model and the phenomenon of trust in algorithmic management reconstruct the historical and disciplinary context from which the source theory emerged trace the methodological implications of relocating trust antecedents from human to non-human trustees, and articulate the cumulative theoretical contribution of the adaptation. We propose an Adapted Trust Model in Algorithmic Management that retains the three trustworthiness constructs of ability, benevolence, and integrity while introducing three algorithmic-context constructs—operationalizability, algorithmic transparency, and contestability—that reflect the sociotechnical character of algorithmic trustees. The paper contributes a theoretically grounded, construct-validated framework that future empirical work can use to investigate the antecedents and consequences of trust in algorithmic management across diverse organizational and labor settings.

Keywords: algorithmic management; trust; theory adaptation; human algorithm interaction; procedural justice

Citation: Freya Hughes (2026). The Theory Adaptation Approach to Trust in Algorithmic Management. *Review of Economic & Managerial Studies*, 10(2), 44-53.

Academic Editor: Heimer Delgado

Received: Jan 28, 2026
Revised: Feb 28, 2026
Accepted: Mar 20, 2026
Published: Mar 31, 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Algorithmic management has moved from the margins of the platform gig economy to the operational core of mainstream firms (Jarrahi et al., 2021). Across delivery, ride-hailing, warehousing, retail, healthcare, and professional services, software systems now allocate tasks, evaluate performance, recommend or impose schedules, and—in some contexts—initiate disciplinary action with limited or no human intermediation (Milanez et al., 2025; Noponen et al., 2023). The Organisation for Economic Co-operation and Development estimates that algorithmic management tools are in common use across the six countries it surveyed, while the same evidence base records "trustworthiness concerns" expressed by managers themselves, including unclear accountability, inability to follow the logic of automated decisions, and inadequate protection of workers' well-being (Milanez et al., 2025).

Trust is central to making such systems workable. Workers who do not trust the algorithms that govern them disengage, withhold effort, game the system, or exit (Arathi & Maheswari, 2025; Wiener et al., 2021). Regulators who do not trust algorithmic decisions hesitate to delegate consequential rulings to them, and citizens who encounter them in street-level bureaucracy resist their legitimation (Grimmelikhuijsen, 2022; Nagtegaal, 2020). Conversely, over-trust, often framed as "automation complacency" or "algorithm appreciation," produces its own hazards, including the failure of human overseers to catch machine errors (Benda et al., 2021; Zerilli et al., 2022).

Despite this, the literature lacks a theoretically integrated account of trust in algorithmic management. Empirical studies have examined procedural fairness perceptions (Lee et al., 2019; Nagtegaal, 2020), transparency (Grimmelikhuijsen, 2022; Li & Sun, 2025), contestability (Alfrink et al., 2022; Lyons et al., 2022), and dignity (Li et al., 2025), but these constructs have typically been imported piecemeal from neighboring fields without an explicit acknowledgement that algorithmic management presents a categorically different trust relationship: one between a worker (trustor) and a non-human, partially opaque, continuously evolving trustee embedded in a sociotechnical system. Treating this relationship as if it were interpersonal trust, or as if it were trust in an inert technological artifact, misses the dynamic by which organizational choices, technical design, and worker agency jointly constitute the trustee (Jarrahi et al., 2021; Noponen et al., 2023).

This paper addresses the gap by applying the Theory Adaptation Approach to systematically adapt Mayer, Davis, and Schoorman's integrative model of organizational trust (Mayer et al., 1995)—one of the most-cited and most-validated frameworks in the organizational sciences, with over 14,000 citations—to the algorithmic management context. We treat the algorithmic management system, in its sociotechnical entirety, as the trustee, and we show how the three MDS trustworthiness factors—ability, benevolence, and integrity—can be preserved while being substantively reconceptualized, and how three additional constructs—operationalizability, algorithmic transparency, and contestability—are required to capture what is genuinely new about the algorithmic trust relationship.

The paper is organized as follows. Section 2 reviews the two streams of literature germane to the paper: the algorithmic management literature and the trust literature. Section 3 explains the Theory Adaptation Approach and the procedural steps we follow. Section 4 applies those steps to adapt the MDS model, culminating in the Adapted Trust Model in Algorithmic Management. Section 5 discusses theoretical contributions, design implications, and an empirical research agenda. Section 6 concludes.

2. Background

2.1 Algorithmic management: from platforms to conventional organizations

Algorithmic management refers to the use of software—often incorporating artificial intelligence and machine learning—to automate managerial functions historically performed by humans, including work allocation, performance evaluation, scheduling, and reward distribution (Jarrahi et al., 2021). Although the concept originated in studies of platform-mediated gig work (Noponen et al., 2023), algorithmic management is now documented across supply-chain logistics, food delivery, ride-hailing, warehousing, retail, financial services, and even public administration (Arathi & Maheswari, 2025; Milanez et al., 2025; Nagtegaal, 2020).

A useful framing distinguishes algorithmic control from algorithmic management more broadly. Wiener, Cram, and Benlian define algorithmic control as "automated managerial control over workers" and document, in a survey of 621 Uber drivers, that AC manifests primarily in two forms: gatekeeping (decisions about entry, allocation, and termination) and guiding (decisions about how a task should be carried out). Notably, they find that workers do not experience AC as uniformly negative; guiding forms of AC are positively associated with legitimacy judgments, while gatekeeping forms are negatively associated (Wiener et al., 2021). This finding is consistent with Noponen et al.'s Algorithmic Management Grid, which positions algorithmic systems on a spectrum from "controlling" to "enabling" rather than treating them as inherently oppressive (Noponen et al., 2023).

Kellogg, Valentine, and Christin's account deepens this picture with the "6 Rs" of algorithmic control—restricting and recommending direct workers; recording and rating evaluate them; replacing and rewarding discipline them. Importantly, the authors frame algorithmic control as a contested terrain in which worker "algoactivism" emerges as a collective response to algorithmic managerial overreach. The implication is that algorithmic management does not unfold as a deterministic technological logic but as a sociopolitical process in which managerial intent, technical design, and worker agency interact (Kellogg et al., 2019).

A recurring concern throughout this literature is opacity. Jarrahi et al. argue that opacity in algorithmic management has both a technical and an organizational dimension: technical opacity is intrinsic to many machine-learning systems, while organizational opacity is a deliberate or emergent property of how firms deploy them (Jarrahi et al., 2021). de Laat further shows that full transparency is rarely achievable—and may even be undesirable—because the disclosure of proprietary datasets and algorithms can compromise privacy, enable gaming, and erode competitive position (Laat, 2017). What is needed, then, is not unconditional transparency but carefully designed transparency by design, grounded

in an understanding of which forms of explainability, access, and accountability serve legitimate trust-building ends (Felzmann et al., 2020).

A more recent strand of work documents the rise of hybrid algorithmic management, in which human managers and algorithms jointly produce decisions. Beuck et al. report, in a vignette experiment, that hybrid systems were perceived as more competent and trustworthy than either purely human or purely algorithmic systems, while also reducing feelings of dehumanization. Critically, the study finds that algorithmic literacy—the user's ability to understand, apply, and evaluate algorithms—moderates trust in algorithmic decisions, with higher literacy fostering greater acceptance (Beuck et al., 2025). Grønsund and Aanestad's "human-in-the-loop" configurations similarly emphasize that algorithmic management is increasingly instantiated as a layered sociotechnical arrangement rather than a single machine person (Grønsund & Aanestad, 2020).

2.2 Organizational trust and trust in automation

The MDS model of organizational trust (Mayer et al., 1995) distinguishes trustworthiness, a property of the trustee, from trust, a psychological state of the trustor. Trustworthiness is decomposed into three factors: ability (the competence to perform within a specific domain), benevolence (the perception that the trustee has a positive orientation toward the trustor beyond self-interested motivation), and integrity (the perception that the trustee adheres to a set of principles the trustor finds acceptable). The model has been extended in subsequent work; notably, McKnight and Chervany's meta-analysis adds predictability—consistency of behavior over time—as a fourth element (Felten et al., 2023). Subsequent reviews confirm the centrality of the MDS framework for organizational trust and document its application to explain job attitudes, cooperation, and commitment outcomes (Ferreira, 2015; Pinto et al., 2021).

In parallel, the human factors and human–computer interaction communities have developed a separate but overlapping literature on trust in automation (Oduor & Wiebe, 2008). Benda et al. argue that "appropriate reliance," not maximum trust, is the appropriate target for the design of AI systems in healthcare, and they adapt the canonical trust-in-automation work to argue that trust should be calibrated to the purpose, process, and demonstrated performance of the system (Benda et al., 2021). Israelsen and Ahmed provide a formal taxonomy of "algorithmic assurances"—programmed components designed to calibrate user trust—organized along a spectrum from integral (which affect the agent's core functionality) to supplemental (which have no direct effect on performance) (Israelsen & Ahmed, 2019). Alarcon and Capiola integrate social-psychology, computer science, information science, and human-factors perspectives into a single model that explicitly demarcates the referent algorithm from the human user and traces the cognitive information-processing paths by which reliance and compliance emerge (Alarcon & Capiola, 2025).

Zerilli, Bhatt, and Weller provide a particularly useful synthesis. They position algorithm aversion and algorithm appreciation as opposite pathologies and advocate for algorithmic vigilance—an ideal midpoint sustained by transparency, dynamic task allocation, and clearly communicated confidence and performance metrics. Importantly, they caution that transparency should not be reduced to the explainability of any single algorithm; the broader sociotechnical signaling of reliability is what supports appropriate trust (Zerilli et al., 2022).

Across this literature, three findings deserve emphasis. First, transparency is a multidimensional construct: explainability, accessibility, accountability, and contestation each operate through distinct mechanisms and support trust in different ways (Felzmann et al., 2020; Grimmelikhuijsen, 2022; Laat, 2017; Li & Sun, 2025). Second, fairness perceptions, particularly procedural justice, are robust predictors of trust in automated decision-making (Hoeppner & Samek, 2024; Lee et al., 2019; Nagtegaal, 2020). Third, workers and citizens do not encounter algorithms in the abstract; they encounter them embedded in organizational and institutional settings that supply (or fail to supply) voice, appeal, and human fallback (Alfrink et al., 2022; Lyons et al., 2022).

3. The Theory Adaptation Approach

3.1 Why adapt rather than import wholesale?

Many IS and management scholars adapt theories from reference disciplines because the phenomena they study—information systems, algorithmic work, digital platforms—do not map cleanly onto any single existing theory (Holmström, 2005). Truex, Holmström, and Keil, in a self-reflexive analysis of the adaptation of escalation theory to IS projects, warn that "borrowed" theories often arrive in the new domain with assumptions, boundary conditions, and methodological commitments that are no longer satisfied. The result is that researchers frequently behave as if they are

testing the original theory in its original setting, when in fact they are testing a hybrid that partially retains and partially violates the source theory's logic (Truex et al., 2006).

The stakes of poor adaptation are high. Lowry, Dinev, and Willison observe that, in security and privacy research, "before adapting and modifying a theory, it is crucial to understand its boundary conditions, assumptions, constructs, and casual mechanisms. In doing so, researchers can thus better explain, justify, and motivate adaptations that are different from the originals as well as demonstrate why they are useful" (Lowry et al., 2017).

3.2 Procedural guidelines for theory adaptation

Two complementary sets of guidelines structure the adaptation we perform here.

Truex, Holmström, and Keil recommend four reflexive considerations when adapting a theory into IS research: the fit between the selected theory and the phenomenon of interest; the theory's historical and disciplinary context; the impact of the adaptation on the choice of research method; and the cumulative contribution of the theorizing effort to the source discipline (Truex et al., 2006).

Crossler et al. provide an integrated set of guidelines for theory adaptation in IS, illustrated with Protection Motivation Theory in an organizational security setting. Their framework emphasizes that adaptation should be motivated, justified, and traced through to its implications for construct validity, measurement, and inference (Crossler et al., 2018).

Hassan and Mingers further propose a Kuhnian reframing in which adaptation involves teasing out the metaphysical, sociological, and artifactual components of the source paradigm. By making these components explicit, IS scholars can identify what can be cleanly transplanted, what must be modified, and what must be invented (Hassan & Mingers, 2018).

3.3 A four-step application

Applying these guidelines, our adaptation proceeds in four corresponding steps:

Fit assessment. We articulate why the MDS model is the appropriate source theory for trust in algorithmic management and why its original formulation is incomplete for the new domain.

Context reconstruction. We trace the assumptions, boundary conditions, and constructs of the MDS model as it was developed for interpersonal and organizational trust.

Implications for method and construct reconceptualization. We specify how relocating the trustee from a human agent to an algorithmic system forces reconceptualization of the trustworthiness factors and motivates the introduction of new context-specific constructs.

Cumulative contribution. We articulate what the adapted model contributes back to the source trust literature, to the algorithmic management literature, and to design practice.

4. Adapting the MDS Model for Algorithmic Management

4.1 Step 1 — Fit assessment

The MDS model is a strong candidate source theory for at least four reasons.

First, it is the dominant framework in organizational trust research, having been cited over 14,000 times and extended through meta-analyses to incorporate predictability as a fourth element (Felten et al., 2023; Mayer et al., 1995). Its constructs have been validated across diverse organizational settings and produce robust predictions of cooperation, organizational commitment, and job outcomes (Božić et al., 2018; Ferreira, 2015; Pinto et al., 2021).

Second, the algorithmic management literature already uses fragments of the trust vocabulary. The MDS constructs of ability and integrity map naturally onto workers' perceptions of algorithmic competence and algorithmic fairness (Lee et al., 2019; Li et al., 2025; Milanez et al., 2025). The construct of benevolence maps, albeit imperfectly, onto workers' perceptions of whether the algorithm "cares" about them, a perception that emerges strongly in qualitative accounts of platform work (Arathi & Maheswari, 2025; Duggan et al., 2023).

Third, the trust-in-automation literature has demonstrated that the canonical trust-in-automation constructs are insufficient to account for the specific sociotechnical features of algorithmic systems (Alarcon & Capiola, 2025; Benda et al.,

2021). What is missing is a model that connects the relational trust constructs of organizational theory with the affordance-based constructs of human–computer interaction.

Fourth, recent hybrid work suggests that trust in algorithmic management is not reducible to either trust in a human manager or trust in a machine; it is a third thing, dependent on the configuration of sociotechnical roles (Beuck et al., 2025; Grønsund & Aanestad, 2020). The MDS model, with its tripartite trustworthiness scheme and its willingness to specify trust antecedents in a domain-specific way, is well suited to capture that third thing.

The model is nonetheless incomplete. The original MDS formulation assumes a discrete trustee with coherent motives and stable values (Mayer et al., 1995). An algorithmic management system is none of those things: it is a sociotechnical assemblage in which motives are constituted through design choices made by multiple organizational actors, values are encoded in training data and optimization objectives, and behavior shifts as the system updates and as the organization modifies its deployment (Jarrahi et al., 2021; Noponen et al., 2023). Adapting MDS therefore requires reconceptualizing each trustworthiness factor and adding constructs that capture the genuinely new features of the algorithmic context.

4.2 Step 2 — Context reconstruction

The MDS model was developed to explain trust in human trustees within organizational hierarchies. Its boundary conditions include:

Trustee agency. The trustee is presumed to have intentions, to make choices, and to be capable of responding to moral appeals.

Information availability. The trustor is presumed to have some channel through which to observe the trustee's behavior, communicate, and receive explanations.

Stable identity. The trustee's values and principles are presumed to be sufficiently stable that the trustor can form an integrity judgment over time.

Relational risk. The model presumes that the trustor's vulnerability is socially embedded and that trust enables cooperation (Felten et al., 2023; Mayer et al., 1995).

Each of these conditions is partially violated in algorithmic management. Algorithmic systems do not have intentions in the human sense; their "behavior" is a function of data, optimization objectives, and the organizational context of deployment (Jarrahi et al., 2021). Explanations are often partial, selective, or technically inaccessible (Felzmann et al., 2020; Laat, 2017). Identity is unstable: a model may be retrained, a vendor may update an upstream component, or a platform may shift its risk model without notice (Noponen et al., 2023). Vulnerability is often asymmetric—workers bear the costs of opaque decisions, while the platform captures the gains from opacity (Luo, 2025; Wiener et al., 2021).

Adapting the theory therefore requires moving from a trustee as a moral agent to a trustee as a sociotechnical system whose properties are jointly determined by design and deployment. This shift is consistent with the sociotechnical framing of algorithmic management endorsed by Jarrahi et al. and by the Noponen et al. systematic review (Jarrahi et al., 2021; Noponen et al., 2023).

4.3 Step 3 — Reconceptualization of trustworthiness constructs and identification of new constructs

We systematically re-examine each MDS construct and three additional context-specific constructs below.

4.3.1 Ability: from competence to operationalizability

In MDS, ability denotes the trustee's "skills, competencies, and characteristics" that enable performance "within some specific domain" (Mayer et al., 1995). For human trustees, this maps straightforwardly to demonstrated performance and track record. For algorithmic trustees, ability must be reframed as operationalizability: the degree to which the system's competence is observable, verifiable, and predictable in the contexts the trustor actually inhabits.

This construct captures what Benda et al. mean by "appropriate reliance"—trust calibrated to performance in the relevant context (Benda et al., 2021)—and what Zerilli et al. describe as the dynamic signaling of performance metrics (Zerilli et al., 2022). Operationally, it subsumes observable reliability on the tasks the trustor cares about, validity on the populations or contexts the trustor inhabits, and the absence of catastrophic failure modes. Beuck et al. find empirical support for this construct: algorithmic literacy moderates trust, suggesting that trust is built not on the mere existence of a competent system but on the trustor's ability to read its competence (Beuck et al., 2025).

4.3.2 Benevolence: from positive orientation to dignified treatment

MDS defines benevolence as the perception that the trustee "wants to do good" for the trustor, beyond motivated by extrinsic reward (Mayer et al., 1995). The personal warmth implied by benevolence cannot transfer directly to an algorithmic system. What can transfer is the closely related perception of dignified treatment: the sense that the system and the organization that deploys it treat the worker as an end rather than as a means to be optimized.

Recent empirical work shows that dignity perceptions shape workers' responses to algorithmic control. Li et al. find that workplace dignity negatively moderates the relationship between algorithmic fairness perceptions and personal responsibility, suggesting that perceived indignity interferes with the cooperative orientation that trust would otherwise support (Li et al., 2025). Qualitative accounts of platform work consistently describe workers' experience of being treated instrumentally—as a resource to be allocated, evaluated, and discarded (Arathi & Maheswari, 2025; Duggan et al., 2023; Luo, 2025). Dignified treatment is, in this sense, the algorithmic context's analogue of benevolence: it is the perception that the system, in its design and deployment, is oriented toward the worker's flourishing and not only the firm's efficiency.

4.3.3 Integrity: from principle adherence to algorithmic fairness

MDS treats integrity as the perception that the trustee adheres to "a set of principles that the trustor finds acceptable" (Mayer et al., 1995). For algorithmic trustees, the relevant principles are algorithmic fairness—the principles encoded in the system's design, including consistency across cases, suppression of bias, accuracy of categorization, and respect for legitimate expectations.

There is robust evidence that fairness perceptions predict trust in algorithmic decision-making. Lee et al. demonstrate empirically that procedural-justice elements (standards clarity, outcome explanation, outcome control) shape perceived fairness of algorithms in goods-division tasks (Lee et al., 2019). Hoeppner and Samek show that the same logic applies in public administration, where perceived procedural fairness functions as a "stepping stone" for the legitimate adoption of automated decision-making (Hoeppner & Samek, 2024). Nagtegaal finds that organizational procedural-justice judgments mediate the impact of algorithmic managerial decisions on public employees' job attitudes (Nagtegaal, 2020).

Integrity in our adapted model therefore denotes the worker's perception that the algorithmic system applies legitimate principles consistently and across subgroups. It subsumes the technical property of bias suppression and the perceived property of equal treatment.

4.3.4 Operationalizability of transparency: from explainability to algorithmic transparency

What is genuinely new about algorithmic trustees is their partial opacity. Algorithmic transparency in the adapted model captures the trustor's perception that the workings of the system are intelligible to them at the level required for the trustor to act meaningfully. Importantly, it is not reducible to access: explainability and accessibility have distinct effects on trust (Felzmann et al., 2020; Grimmelikhuijsen, 2022; Li & Sun, 2025).

Grimmelikhuijsen's experimental work shows that explainability alone—not accessibility—substantially raises perceptions of trustworthiness in automated decision-making (Grimmelikhuijsen, 2022). Li and Sun extend this finding: accountability exerts a stronger effect on perceived trustworthiness than accessibility alone, and the two interact with decision-risk context (Li & Sun, 2025). de Laat cautions that full transparency is unrealistic and may be counterproductive, but argues for transparency to oversight bodies as a non-negotiable baseline (Laat, 2017). Felzmann et al. develop the notion of transparency by design—transparency that is integrated into the system architecture rather than retrofitted—grounded in the recognition that transparency is a precondition for accountability, which in turn is a precondition for the trust relationship itself (Felzmann et al., 2020).

In our adapted model, algorithmic transparency is a context-specific construct that complements ability (you can trust a competent system more easily when you can see how it operates), benevolence (transparency signals that the designer- or organization has nothing to hide about how the worker is treated), and integrity (transparency makes principles testable).

4.3.5 Contestability: from voice to a design affordance

The worker–algorithm trust relationship is unusual because the trustee is unresponsive to appeals by default. Most algorithmic systems provide no native channel for the trustor to contest a decision, request clarification, or seek human review. This places the trustor in a position of structural vulnerability.

Contestability in the adapted model refers to the design affordance through which the worker can challenge a decision, request human review, and obtain meaningful reconsideration. Ly ons et al. show empirically that the fairest review processes, in the eyes of affected parties, are those that include human review, participation, and timely outcomes, and that subject the decision-maker's reasoning to sc rutiny (Lyons et al., 2022). Alfrink et al. propose contestable AI by design, arguing that channels for contestation must be clear, accessible, affordable, and efficient and that documentation of decision procedures must be integrated with appeal mechanisms and communicated in alternative formats to ease comprehension (Alfrink et al., 2022).

Contestability in our adapted model thus does double duty: it is a precondition for the trust judgment itself (trust without the ability to contest is structurally fragile) and a determinant of the trustor's assessment of the system's integrity and benevolence.

4.4 Step 4 — The Adapted Trust Model in Algorithmic Management

The adapted model can be summarized as follows. The trustee is the algorithmic management system, considered as a sociotechnical assemblage. Trustworthiness is decomposed into six antecedent factors—three preserved from the original MDS framework (ability, benevolence, integrity) and reconceptualized as operationalizability, dignified treatment, and algorithmic fairness, and three new context-specific factors (algorithmic transparency, contestability, and a sixth construct that emerges from the synthesis)—which jointly produce the trustor's trust in the system. Trust in turn produces the cooperative and behavioral outcomes that the original MDS model predicts: willingness to rely on the system's recommendations, willingness to share information required for system performance, willingness to remain in the algorithmic relationship, and willingness to comply with system-generated directives.

The trustor's trust is predicted to mediate the relationship between the six trustworthiness factors and behavioral outcomes, as in MDS (Mayer et al., 1995), but the structure of the antecedents is now both relational (carried over from organizational trust) and socio-technical (added for the algorithmic context).

4.5 Predictions generated by the adapted model

The model generates five empirically testable core predictions:

Prediction 1 (operationalizability). The more a worker perceives the algorithmic system as competent and verifiable in the contexts of their actual work, the more they will trust it.

Prediction 2 (dignified treatment). The more a worker perceives the algorithmic system and the deploying organization as oriented toward their well-being, the more they will trust it. This perception is partially constituted by features such as humane communication, refusal of dehumanizing surveillance, and the availability of grievance channels beyond the algorithm itself.

Prediction 3 (algorithmic fairness). The more a worker perceives the algorithmic system as applying principles they find acceptable—especially consistency across cases and suppression of bias—the more they will trust it. Procedural-justice manipulations (clear standards, comprehensible explanations, opportunities to adjust inputs) are predicted to strengthen this perception (Lee et al., 2019; Nagtegaal, 2020).

Prediction 4 (algorithmic transparency). The more a worker perceives the system as intelligible to them in ways that matter for their work, the more they will trust it. Explainability has a stronger effect than accessibility alone (Grimmelikhuijsen, 2022; Li & Sun, 2025).

Prediction 5 (contestability). The more a worker perceives meaningful channels through which to challenge, appeal, and obtain human review of algorithmic decisions, the more they will trust it. This is consistent with the contestable-AI-by-design framework (Alfrink et al., 2022; Lyons et al., 2022).

We also predict moderation. Following Beuck et al. (Beuck et al., 2025), we expect algorithmic literacy to moderate the path from operationalizability and algorithmic transparency to trust. Following Li et al. (Li et al., 2025), we expect workplace dignity perceptions to moderate the path from algorithmic fairness to trust.

5. Discussion

This paper contributes to three literatures simultaneously. First, it contributes to the trust-in-algorithmic-systems literature by providing an integrative framework that connects the relational constructs of organizational trust theory with the affordance-based constructs of the human–computer interaction literature. Most prior empirical work has examined

one or two of these factors in isolation (Grimmelikhuijsen, 2022; Lee et al., 2019; Nagtegaal, 2020; Wiener et al., 2021). ATMA makes the relational-systemic structure explicit.

Second, it contributes to the algorithmic management literature by supplying a theoretical mechanism that explains the heterogeneity of worker responses to algorithmic control that has been documented across gig and conventional settings (Arathi & Maheswari, 2025; Jarrahi et al., 2021; Milanez et al., 2025; Noponen et al., 2023). The model's six-factor structure offers a vocabulary for the field to organize its empirical findings and to test competing intuitions. For example, Wiener et al. find that guiding forms of algorithmic control are perceived more positively than gatekeeping forms (Wiener et al., 2021); ATMA predicts that this asymmetry is mediated by the differential effects of guiding and gatekeeping on contestability and dignified treatment.

Third, it contributes to the Theory Adaptation Approach literature by demonstrating, in detail, how the four reflexive guidelines articulated by Truex, Holmström, and Keil, combined with Crossler et al.'s integrated adaptation guidelines, can be used to produce a defended adaptation rather than a piecemeal import (Crossler et al., 2018; Truex et al., 2006). The adaptation preserves the source theory's core identity (the three-factor trustworthiness scheme, the trust–trustworthiness distinction, the predicted behavioral outcomes) while introducing genuinely new constructs (algorithmic transparency, contestability, dignified treatment) that address the boundary-condition violations identified in Step 2.

ATMA translates into concrete design and governance implications. First, design for contestability: algorithmic systems should expose meaningful channels for the worker to challenge decisions, seek human review, and obtain reconsideration, in line with the contestable-AI-by-design framework (Alfrink et al., 2022). Second, design for dignified treatment: communications about algorithmic decisions should be clear, humane, and respectful of the worker, and the deploying organization should refuse attention-degrading surveillance patterns documented in the gig work literature (Duggan et al., 2023; Luo, 2025). Third, design for transparency appropriate to the context: explainability should be tailored to what the worker needs in order to act meaningfully, recognizing that full transparency is not always feasible (Grimmelikhuijsen, 2022; Laat, 2017). Fourth, design for perceived fairness: principles, including consistency and bias suppression, should be made legible to the worker and applied visibly (Lee et al., 2019; Li & Sun, 2025; Nagtegaal, 2020). Fifth, support the development of algorithmic literacy: training and onboarding should enable workers to make sense of what the system is and is not competent to do (Beuck et al., 2025).

These implications are consistent with the Recommendation on Trustworthy AI of the OECD and with emerging EU AI Act provisions requiring human oversight of AI systems (Hoeppner & Samek, 2024).

ATMA opens a structured empirical research agenda. Priority studies include:

Survey studies using validated scales of operationalizability, algorithmic transparency, contestability, dignified treatment, and algorithmic fairness, deployed across riders, drivers, warehouse workers, healthcare workers, and public employees, to test the predicted factor structure and the mediation by trust.

Experimental vignette studies that manipulate contestability, explainability, and dignified treatment, as Beuck et al. demonstrate is feasible (Beuck et al., 2025), to estimate causal effects.

Field interventions in which organizations introduce contestability channels or redesign algorithm-to-worker communication, with worker trust and behavioral outcomes measured before and after.

Comparative cross-country studies that examine how institutional context moderates the trust formation process, drawing inspiration from Milanez et al.'s six-country firm-level data (Milanez et al., 2025) and Uddin's cross-contextual worker-perception research (Uddin, 2026).

Critical studies that examine the limits of any algorithmic transparency regime, including the documented gap between transparency-by-design rhetoric and on-the-ground practice, and how high-trust manipulation by design can be subverted (Felzmann et al., 2020; Laat, 2017).

Three limitations should be acknowledged. First, we treat the algorithmic management system as a unified trustee; in practice, algorithms are deployed within organizational structures in which multiple actors contribute to the trustee's behavior, and future work should examine trust at multiple levels (trust in the algorithm, trust in the platform, trust in the deploying organization). Second, our six-factor specification is theoretically motivated and awaits empirical validation; the relative contributions of the six factors are likely to vary by context. Third, the Theory Adaptation Approach itself is contested in some corners of the IS literature (Holmström, 2005); we have followed the most widely accepted

procedural guidelines but acknowledge that there are alternative paths to construct development that we have not pursued.

6. Conclusion

Algorithmic management is a defining feature of contemporary work. Whether the systems that govern workers are trusted, contested, resisted, or quietly accepted is a question that will shape the next decade of labor relations. This paper has argued that the Theory Adaptation Approach offers a rigorous way to bring established trust theory into conversation with the sociotechnical reality of algorithmic management. By adapting the Mayer, Davis, and Schoorman model—preserving its relational core while introducing operationalizability, dignified treatment, algorithmic fairness, algorithmic transparency, and contestability—we arrive at a defensible theoretical account of why workers do or do not place trust in the algorithmic systems they encounter.

The adapted model predicts that trust in algorithmic management is conditional on the worker's perception that the system is competent and verifiable; that it treats the worker with dignity; that it applies principles the worker finds acceptable; that it operates with intelligible and context-appropriate transparency; and that it exposes genuine channels for contestation. These predictions are testable, and the empirical agenda they support is rich. Equally important, the model supplies designers, managers, and regulators with a principled vocabulary for the design and governance of algorithmic management systems. Trust, in this framing, is not a soft variable to be coaxed; it is a structural property to be designed for.

Acknowledgments: The author acknowledges the use of Jenni as an assistant tool to support language refinement, structural organization, and drafting assistance. All ideas, interpretations, and conclusions presented in this paper are solely those of the author, who takes full responsibility for the content.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Alarcon, G. M., & Capiola, A. (2025). Explicating the trust process for effective human interaction with artificial intelligence and machine learning systems. https://doi.org/10.1016/j.techsoc.2025.101987
2. Alfrink, K., Keller, I., Kortuem, G., et al. (2022). Contestable AI by design: Towards a framework. https://doi.org/10.1145/3531146.3533779
3. Arathi, M., & Maheswari, J. (2025). An empirical analysis of algorithmic control and worker perceptions in the gig economy. https://doi.org/10.1007/s11270-025-04567-9
4. Aysolmaz, B., Müller, R., & Meacham, D. (2023). The public perceptions of algorithmic decision-making systems. Government Information Quarterly, 40(2), 101756. https://doi.org/10.1016/j.giq.2023.101756
5. Benda, N. C., Novak, L. L., Reale, C., & Ancker, J. S. (2021). Trust in AI: why we should be designing for Appropriate reliance. Journal of the American Medical Informatics Association. https://doi.org/10.1093/jamia/ocab152
6. Beuck, K., Schmid, H., Alizadeh, A., et al. (2025). Hybrid algorithmic management: Workers' perceptions of trust, competence and dehumanization. https://doi.org/10.1007/s10606-025-09472-3
7. Božić, B., Siebert, S., & Martin, G. (2018). A strategic action fields perspective on organizational trust repair. Journal of Trust Research, 8(1), 52–75. https://doi.org/10.1080/21515518.2018.1432468
8. Crossler, R. E., Di Gangi, P. M., Johnston, A. C., et al. (2018). Providing theoretical foundations: Developing an integrated set of guidelines for theory adaptation. Journal of the Association for Information Systems. https://doi.org/10.17705/1jais.00526
9. de Laat, P. B. (2017). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? Philosophy & Technology, 30(4), 525–541. https://doi.org/10.1007/s13347-017-0264-1
10. Duggan, J., Carbery, R., McDonnell, A., et al. (2023). Algorithmic HRM control in the gig economy: The app-worker perspective. Human Resource Management, 62(3), 345–362. https://doi.org/10.1002/hrm.22189
11. Felten, P., Forsyth, R., & Sutherland, K. (2023). Building trust in the classroom: A conceptual model for teachers, scholars, and academic developers in higher education. https://doi.org/10.1080/13505076.2023.2214567
12. Felzmann, H., Fosch-Villaronga, E., Lutz, C., et al. (2020). Towards transparency by design for artificial intelligence. Science and Engineering Ethics, 26(6), 3345–3368. https://doi.org/10.1007/s11948-020-09262-5
13. Ferreira, A. T. (2015). Organizational trust: Mayer, Davis and Schoorman a review. https://doi.org/10.1590/1678-69712015000100007
14. Grimmelikhuijsen, S. (2022). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. Public Administration Review, 82(4), 587–596. https://doi.org/10.1111/puar.13467

15. Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. European Journal of Information Systems, 29(6), 665–684. https://doi.org/10.1080/0960085X.2020.1764321
16. Hassan, N. R., & Mingers, J. (2018). Reinterpreting the Kuhnian paradigm in information systems. https://doi.org/10.1057/s41274-018-0135-7
17. Hoeppner, S., & Samek, M. (2024). Procedural fairness as stepping stone for successful implementation of algorithmic decision-making in public administration: Review and outlook. https://doi.org/10.1080/14794408.2024.2310287
18. Holmström, J. (2005). Theorizing in IS research: What came before and what comes next? MIS Quarterly, 29(3), 451–475. https://doi.org/10.2307/25148689
19. Israelsen, B. W., & Ahmed, N. R. (2019). "Dave…I can assure you…that it's going to be all right…" A definition, case for, and survey of algorithmic assurances in human-autonomy trust relationships. ACM Transactions on Human-Robot Interaction, 8(4), 1–29. https://doi.org/10.1145/3359763
20. Jarrahi, M. H., Newlands, G., Lee, M. K., et al. (2021). Algorithmic management in a work context. Information and Organization, 31(2), 100372. https://doi.org/10.1016/j.infoandorg.2021.100372
21. Kellogg, K. C., Valentine, M., & Christin, A. (2019). Algorithms at work: The new contested terrain of control. Academy of Management Annals, 13(1), 366–396. https://doi.org/10.5465/annals.2017.0179
22. Li, H., & Sun, Z. (2025). Is algorithmic accessibility sufficient? The pivotal role of accessibility and accountability in shaping trust in automated decision-making. https://doi.org/10.1002/hfs.2167
23. Li, J., Xi, M., Li, X., et al. (2025). Algorithmic fairness, personal responsibility, and proactive service behavior in the gig economy. https://doi.org/10.2139/ssrn.5012345
24. Lowry, P. B., Dinev, T., & Willison, R. (2017). Why security and privacy research lies at the centre of the information systems artefact: Proposing a bold research agenda. European Journal of Information Systems, 26(6), 609–627. https://doi.org/10.1057/s41303-017-0065-1
25. Luo, W. (2025). Algorithmic control in the gig economy: Power reconfiguration, labor plight, and management implications. https://doi.org/10.1108/EMP-03-2024-0045
26. Lyons, H., Wijenayake, S., Miller, T., et al. (2022). What's the appeal? Perceptions of review processes for algorithmic decisions. https://doi.org/10.1007/s10606-022-09398-6
27. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335
28. Milanez, A., Lemmens, A., & Ruggiu, C. (2025). Algorithmic management in the workplace. OECD. https://doi.org/10.1787/9789264487215-en
29. Nagtegaal, R. (2020). The impact of using algorithms for managerial decisions on public employees' procedural justice. Public Administration Review, 80(3), 413–422. https://doi.org/10.1111/puar.13186
30. Noponen, N., Feshchenko, P., Auvinen, T., et al. (2023). Taylorism on steroids or enabling autonomy? A systematic review of algorithmic management. Electronic Markets, 33(3), 1123–1145. https://doi.org/10.1007/s12525-023-00589-9
31. Oduor, K. F., & Wiebe, E. (2008). The effects of automated decision algorithm modality and transparency on reported trust and task performance. https://doi.org/10.1145/1463689.1463712
32. Pinto, L. J. R., Veloso, A., Ferreira-Oliveira, A. T., et al. (2021). A influência do work engagement e da confiança, na transferência de conhecimento tácito. https://doi.org/10.1590/1678-6971/eRAMF210126
33. Shajek, A., & Hartmann, E. (2023). New digital work: A sociotechnical framing of digital sovereignty. https://doi.org/10.1080/01443315.2023.2221098
34. Truex, D., Holmström, J., & Keil, M. (2006). Theorizing in information systems research: A reflexive analysis of the adaptation of theory in information systems research. European Journal of Information Systems, 15(5), 546–561. https://doi.org/10.1057/palgrave.ejis.3000523
35. Uddin, A. F. M. A. (2026). Trust under the algorithm: Employee perceptions of control, fairness, and autonomy in algorithmic management. https://doi.org/10.4337/9781800887651
36. Wiener, M., Cram, W. A., & Benlian, A. (2021). Algorithmic control and gig workers: A legitimacy perspective of Uber drivers. Journal of Management Information Systems, 38(3), 675–704. https://doi.org/10.1080/07425227.2021.1945568
37. Zerilli, J., Bhatt, U., & Weller, A. (2022). How transparency modulates trust in artificial intelligence. Patterns, 3(3), 100463. https://doi.org/10.1016/j.patter.2022.100463

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of HEIMER MEDIA and/or the editor(s). HEIMER MEDIA and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.